

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2002, 18:45:47 ; Search time 2.90323 Seconds  
(without alignments)  
128.577 Million cell updates/sec

Title: US-09-674-593-3  
Perfect score: 57  
Sequence: 1 LPRWPPQL 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues  
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	44	77.2	1733	1	VNUA_PRVKA	P33485 pseudorabies
2	41	71.9	410	1	KYK2_DICDI	P18161 dictyosteli
3	41	71.9	789	1	K6PF_HAECO	Q27665 haemochus
4	40	70.2	1274	1	ENAM_MOUSE	O55196 mus musculus
5	39	68.4	197	1	Y089_MYCTU	Q10886 mycobacteri
6	39	68.4	280	1	CC2B_ANTWA	Q38773 antirrhinum
7	39	68.4	715	1	S141_HUMAN	Q92503 homo sapien
8	39	68.4	852	1	WS14_HUMAN	Q9np71 homo sapien
9	38	66.7	184	1	T13C_HUMAN	Q96rj3 homo sapien
10	38	66.7	310	1	Y497_MYCTU	Q11162 mycobacteri
11	38	66.7	720	1	AB13_ARATH	Q01593 arabidopsis
12	38	66.7	1242	1	WDRA_HUMAN	Q9hb66 homo sapien
13	38	66.7	2411	1	DAB_DROME	P98081 drosophila
14	37.5	65.8	5035	1	RYR1_PIG	P16960 sus scrofa
15	37.5	65.8	5037	1	RYR1_RABIT	P11716 oryctolagus
16	37	64.9	13	1	TP13_PHYTO	P04096 phyllomedusa
17	37	64.9	194	1	HLV2_ACTPL	P46028 actinobacil
18	37	64.9	265	1	COX3_MYTED	P41775 mytilus edu
19	37	64.9	294	1	CC2_VIGUN	P52389 vigna ungui
20	37	64.9	340	1	MPCP_CAEEL	P40614 caenorhabdi
21	37	64.9	356	1	MPCP_RAT	P16036 rattus norv
22	37	64.9	357	1	MPCP_MOUSE	Q8vem8 mus musculu
23	37	64.9	362	1	MPCP_BOVIN	P12234 bos taurus
24	37	64.9	362	1	MPCP_HUMAN	Q00325 homo sapien
25	37	64.9	369	1	CYRG_MOUSE	P34902 mus musculu
26	37	64.9	405	1	FUT4_HUMAN	P22083 homo sapien
27	37	64.9	433	1	FUT4_RAT	Q62994 rattus norv
28	37	64.9	485	1	SOC7_HUMAN	O14512 homo sapien
29	37	64.9	486	1	GATB_AERPE	Q9yb79 aeropyrum p
30	37	64.9	509	1	RA18_MOUSE	Q9qkx2 mus musculu
31	37	64.9	515	1	ENV_BLV	P51519 bovine leuk
32	37	64.9	515	1	ENV_BLVAF	P25504 bovine leuk
33	37	64.9	515	1	ENV_BLVAV	P25057 bovine leuk

34	37	64.9	515	1	ENV_BLVB2	P25506 bovine leuk
35	37	64.9	515	1	ENV_BLVBS	P25507 bovine leuk
36	37	64.9	515	1	ENV_BLVJ	P03380 bovine leuk
37	37	64.9	536	1	C7C2_MAIZE	Q43255 zea mays (m
38	37	64.9	614	1	ACES_RAT	P37136 rattus norv
39	37	64.9	767	1	ACES_CHICK	P36196 gallus gall
40	37	64.9	1001	1	PTPX_MOUSE	P80560 mus musculu
41	37	64.9	1130	1	C2TA_HUMAN	P33076 homo sapien
42	37	64.9	1173	1	GIGA_ORYSA	Q9awl7 oryza sativ
43	37	64.9	1234	1	PIP3_MOUSE	P51432 mus musculu
44	37	64.9	1395	1	CUT1_MOUSE	P53564 mus musculu
45	37	64.9	1783	1	RAA3_CHLRE	Q9fec4 chlamydomon

ALIGNMENTS

RESULT 1  
VNUA\_PRVKA  
ID VNUA\_PRVKA STANDARD; PRT: 1733 AA.  
AC P33485;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 01-FEB-1994 (Rel. 28, Last annotation update)  
DE Probable nuclear antigen.  
OS Pseudorabies virus (strain Kaplan) (PRV).  
OC Viruses: dsDNA viruses, no RNA stage; Herpesviridae;  
OC Alphaherpesvirinae; Varicellovirus.  
OX NCBI\_TaxID=33703;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91021039; PubMed=2171211;  
RA Vleck C., Kozmik Z., Paces V., Schirm S., Schwyzler M.;  
RT "Pseudorabies virus immediate-early gene overlaps with an oppositely  
oriented open reading frame: characterization of their promoter and  
enhancer regions.";  
RL Virology 179:365-377(1990).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M34651; AAA47471.1; -  
DR PIR; B45344; B45344.  
FT DOMAIN 112 117 POLY-THR.  
FT DOMAIN 179 1733 GLY-RICH.  
FT DOMAIN 192 196 POLY-SER.  
FT DOMAIN 271 298 POLY-PRO.  
FT DOMAIN 304 308 POLY-ARG.  
FT DOMAIN 883 889 POLY-GLY.  
FT DOMAIN 1398 1405 POLY-GLY.  
SQ SEQUENCE 1733 AA; 172166 MW; 0C8CD8BE475BB5E2 CRC64;

Query Match 77.2%; Score 44; DB 1; Length 1733;  
Best Local Similarity 100.0%; Pred. No. 47;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 PRWPPP 7  
Db 213 PRWPPP 218

RESULT 2  
KYK2\_DICDI  
ID KYK2\_DICDI STANDARD; PRT: 410 AA.  
AC P18161;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)

```
DE Tyrosine-protein kinase 2 (EC 2.7.1.112) (Fragment).
GN PYKB OR DPYK2.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90287147; PubMed=1972546;
RA Tan J.L., Spudich J.A.;
RT "Developmentally regulated protein-tyrosine kinase genes in
RT Dictyostelium discoideum."
RL Mol. Cell. Biol. 10:3578-3583(1990).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M33784; AAA33203.1; -
DR PIR; B35670; B35670.
DR HSP; P08631; IAD5.
DR DictyDb; D003011; pyKB.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR004040; STY_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00221; STYK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT NON_TER 1
FT DOMAIN 108 381 PROTEIN KINASE.
FT NP_BIND 114 122 ATP (BY SIMILARITY).
FT BINDING 135 135 ATP (BY SIMILARITY).
FT ACT_SITE 232 232 BY SIMILARITY.
SQ SEQUENCE 410 AA; 46386 MW; E93918B605B9AEC1 CRC64;

Query Match 71.9%; Score 41; DB 1; Length 410;
Best Local Similarity 75.0%; Pred. No. 30;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 PRWPPQL 9
Db 273 PRWRPPEL 280
||| ||:|

RESULT 3
K6PF_HAECO STANDARD; PRT; 789 AA.
ID K6PF_HAECO
AC Q27665;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 6-phosphofructokinase (EC 2.7.1.11) (Phosphofructokinase)
DE (Phosphohexokinase).
GN PFK.
OS Haemochus contortus (Barber pole worm).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
OC Trichostrongyloidea; Haemochidae; Haemochinae; Haemochus.
OX NCBI_TaxID=6289;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92140452; PubMed=1838137;
RA Klein R.D., Faureau M.A., Winterrowd C.A., Hatzenbuehler N.T.,
RA Shea M.H., Nulif S.C., Geary T.G., Olson E.R.;

Tyrosine-protein kinase 2 (EC 2.7.1.112) (Fragment).
GN PYKB OR DPYK2.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90287147; PubMed=1972546;
RA Tan J.L., Spudich J.A.;
RT "Developmentally regulated protein-tyrosine kinase genes in
RT Dictyostelium discoideum."
RL Mol. Cell. Biol. 10:3578-3583(1990).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M33784; AAA33203.1; -
DR PIR; B35670; B35670.
DR HSP; P08631; IAD5.
DR DictyDb; D003011; pyKB.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR004040; STY_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00221; STYK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT NON_TER 1
FT DOMAIN 108 381 PROTEIN KINASE.
FT NP_BIND 114 122 ATP (BY SIMILARITY).
FT BINDING 135 135 ATP (BY SIMILARITY).
FT ACT_SITE 232 232 BY SIMILARITY.
SQ SEQUENCE 410 AA; 46386 MW; E93918B605B9AEC1 CRC64;

Query Match 71.9%; Score 41; DB 1; Length 410;
Best Local Similarity 75.0%; Pred. No. 30;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 PRWPPQL 9
Db 273 PRWRPPEL 280
||| ||:|

RESULT 3
K6PF_HAECO STANDARD; PRT; 789 AA.
ID K6PF_HAECO
AC Q27665;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 6-phosphofructokinase (EC 2.7.1.11) (Phosphofructokinase)
DE (Phosphohexokinase).
GN PFK.
OS Haemochus contortus (Barber pole worm).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
OC Trichostrongyloidea; Haemochidae; Haemochinae; Haemochus.
OX NCBI_TaxID=6289;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92140452; PubMed=1838137;
RA Klein R.D., Faureau M.A., Winterrowd C.A., Hatzenbuehler N.T.,
RA Shea M.H., Nulif S.C., Geary T.G., Olson E.R.;

"Cloning of a cDNA encoding phosphofructokinase from Haemochus
contortus.";
Mol. Biochem. Parasitol. 48:17-26(1991).
-!- CATALYTIC ACTIVITY: ATP + D-fructose 6-phosphate = ADP + D-
fructose 1,6-bisphosphate.
-!- PATHWAY: KEY CONTROL STEP OF GLYCOLYSIS.
-!- SIMILARITY: BELONGS TO THE PHOSPHOFRUCTOKINASE FAMILY. TWO DOMAIN
SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M59805; AAA29181.1; -
DR HSP; P06998; 2PFK.
DR InterPro; IPR000023; Ppfruckinase.
DR Pfam; PF00365; PFK; 2.
DR PRINTS; PR00476; PPFRTCKINASE.
DR ProDom; PD000707; Ppfruckinase; 2.
DR PROSITE; PS00433; PHOSPHOFRUCTOKINASE; 2.
DR Kinase; Transferase; Glycolysis; Repeat.
KW SEQUENCE 789 AA; 87184 MW; 1245288AE2289BCF CRC64;

Query Match 71.9%; Score 41; DB 1; Length 789;
Best Local Similarity 71.4%; Pred. No. 58;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LPRWPPP 7
Db 246 IPEWPPP 252
|| |||||

RESULT 4
ENAM_MOUSE STANDARD; PRT; 1274 AA.
ID ENAM_MOUSE
AC O55196;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Enamelin precursor.
GN ENAM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-Swiss Webster; TISSUE=Enamel epithelium;
RX MEDLINE=20514993; PubMed=11062988;
RA Hu C.-C., Simmer J.P., Bartlett J.D., Nanci A., Qian Q., Zhang C.,
RA Ryu O.H., Xue J., Fukae M., Uchida T., McDougall M.;
RT "Murine enamel: cDNA and derived protein sequences.";
RL Connect. Tissue Res. 39:47-61(1998).
CC -!- FUNCTION: PEPTIDES DERIVED FROM THE PARENT ENAMELIN ARE COMPONENTS
CC OF ENAMEL, A UNIQUE AND HIGHLY MINERALIZED ECTODERMAL TISSUE
CC COVERING VERTEBRATE TEETH (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U82698; AAB94312.1; -
DR MGD; MGI:133372; Enam.
KW Signal; Enamel; Glycoprotein.
FT SIGNAL 1 38
```

```
FT CHAIN 39 1274 ENAMELIN.
FT CARBOHYD 130 252 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 259 259 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 269 269 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1066 1066 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1274 AA; 140954 MW; F9DBD1CC9D327143 CRC64;

Query Match 70.2%; Score 40; DB 1; Length 1274;
Best Local Similarity 71.4%; Pred. No. 1.3e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LRPWPPP 7
DB 97 MPWPPPP 103

RESULT 5
Y089_MYCTU STANDARD; PRT; 197 AA.
AC Q10886;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative methyltransferase RV0089 (EC 2.1.1.-).
GN RV0089 OR MT0098 OR MTCY251.07.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H378V;
RX MEDLINE=98299987; PubMed=9634230;
RA Gordon S.V., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Cote S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RL complete genome sequence.";
RN Nature 393:537-544(1998).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RL laboratory strains.";
RC Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE METHYLTRANSFERASE SUPERFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z74410; CAA98925.1; -.
CC EMBL; AE006921; AAK44320.1; -.
CC TIGR; MT0098; -.
CC TuberculList; RV0089; -.
CC InterPro; IPR001601; Methyltransf.
CC InterPro; IPR001737; RNA_A_dimeth.
```

```
DR InterPro; IPR000051; SAM_bind.
DR Pfam; PF00398; RnaAD; 1.
KW Hypothetical protein; Transferase; Methyltransferase;
KW Complete proteome.
SQ SEQUENCE 197 AA; 21596 MW; 004334F1A289EC51 CRC64;

Query Match 68.4%; Score 39; DB 1; Length 197;
Best Local Similarity 83.3%; Pred. No. 29;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 RWPPPQ 8
DB 158 KWPPPQ 163

RESULT 6
CC2B_ANTMA STANDARD; PRT; 280 AA.
AC Q38773;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Cell division control protein 2 homolog B (EC 2.7.1.1-) (Fragment).
GN CDC2B.
OS Antirrhinum majus (Garden snapdragon).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Lamiales; Veronicaceae; Antirrhinum.
OX NCBI_TaxID=4151;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Flower;
RX MEDLINE=96434535; PubMed=8837502;
RA Robert P.R., Gaudin V., Lunness P., Coen E.S., Doonan J.H.;
RT "Distinct classes of cdc2-related genes are differentially expressed
RL during the cell division cycle in plants.";
CC Plant Cell 8:1465-1476(1996).
CC -!- FUNCTION: PLAYS A KEY ROLE IN THE CONTROL OF THE EUKARYOTIC
CC CELL CYCLE.
CC -!- ENZYME REGULATION: PHOSPHORYLATION AT TYR-2 INACTIVATES THE
CC ENZYME, WHILE PHOSPHORYLATION AT THR-148 ACTIVATES IT (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC CDC2/CDKX SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X97638; CAA66234.1; -.
CC HSP; P24941; IHCL.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00220; S_TKC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; PARTIAL.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
KW Transferase; Serine/threonine-protein kinase; ATP-binding;
KW Cell cycle; Cell division; Mitosis; Phosphorylation.
FT NON_TER 1 1
FT DOMAIN <1 274 PROTEIN KINASE.
FT NP_BIND <1 5 ATP (BY SIMILARITY).
FT BINDING 20 20 ATP (BY SIMILARITY).
FT ACT_SITE 114 114 BY SIMILARITY.
FT MOD_RES 2 2 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 148 148 PHOSPHORYLATION (BY CAK) (BY SIMILARITY).
SQ SEQUENCE 280 AA; 32167 MW; EB4F3715A4FF66CC CRC64;
```

Query Match 68.4%; Score 39; DB 1; Length 280;  
Best Local Similarity 62.5%; Pred. No. 41;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 PRWPPPOL 9  
|:|:|:|:  
Db 229 PKWPPREL 236

## RESULT 7

S141\_HUMAN STANDARD; PRT: 715 AA.  
AC Q92503; Q99780;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE SEC14-like protein 1.  
GN SEC14L1 OR SEC14L.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lung;  
RX MEDLINE=96302338; PubMed=8697811;  
RA Chinen K., Takahashi E., Nakamura Y.;  
RT "Isolation and mapping of a human gene (SEC14L), partially homologous  
RT to yeast SEC14, that contains a variable number of tandem repeats  
RT (YNT) site in its 3' untranslated region.";  
RL Cytogenet. Cell Genet. 73:218-223(1996).  
RN [2]  
RP SEQUENCE OF 423-715 FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=97264341; PubMed=9110174;  
RA Yu W., Andersson B., Worley K.C., Muzny D.M., Ding Y., Liu W.,  
RA Ricafrente J.Y., Wentland M.A., Lennon G., Gibbs R.A.;  
RT "Large-scale concatenation cDNA sequencing.";  
RL Genome Res. 7:353-358(1997).  
CC -1- TISSUE SPECIFICITY: UBIQUITOUS.  
CC -1- SIMILARITY: CONTAINS 1 CRAL-TRIO DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: D67029; BAAL1048.1; -  
CC EMBL: U79284; AAB50220.1; -  
CC Genew; HGNC:10698; SEC14L1.  
CC MIM; 601504; -  
CC InterPro; IPR001251; CRAL-TRIO.  
CC Pfam; PF00650; SEC14; 1.  
CC SMART; SM00516; CRAL-TRIO; 1.  
CC PROSITE; PS01091; CRAL-TRIO; 1.  
FT DOMAIN 319 495  
FT CONFLICT 715 715 R -> RWRC (IN REF. 2).  
SQ SEQUENCE 715 AA; 81277 MW; F35B08CD8ADDE79A CRC64;

Query Match 68.4%; Score 39; DB 1; Length 715;  
Best Local Similarity 55.8%; Pred. No. 1e+02;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 LPRWPPPOL 9  
:|:|:|:|:  
Db 178 VPRWSPPI 186

## RESULT 8

WS14\_HUMAN STANDARD; PRT: 852 AA.  
ID Q9N7L1; Q9BY06; Q9BY04; Q9BY03; Q96E48; Q9Y2P3;  
AC 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Williams-Beuren syndrome chromosome region 14 protein (WS basic-helix-  
DE loop-helix leucine zipper protein) (WS-BHLH) (Mlx interactor).  
GN WBSCR14 OR MIO.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RX MEDLINE=20241700; PubMed=10780788;  
RA de Luis O., Valero M.C., Perez Jurado L.A.;  
RT "WBSCR14, a putative transcription factor gene deleted in Williams-  
RT Beuren syndrome: complete characterisation of the human gene and the  
RT mouse ortholog.";  
RL Eur. J. Hum. Genet. 8:215-222(2000).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5), AND SUBUNIT.  
RX PubMed=11230181;  
RA Cairo S., Merla G., Urbinati F., Ballabio A., Raymond A.;  
RT "WBSCR14, a gene mapping to the Williams-Beuren syndrome deleted  
RT region, is a new member of the Mlx transcription factor network.";  
RL Hum. Mol. Genet. 10:617-627(2001).  
RN [3]  
RP SEQUENCE OF 620-852 FROM N.A. (ISOFORM 4).  
RX MEDLINE=99075645; PubMed=9860302;  
RA Meng X., Lu X., Li Z., Green E.D., Massa H., Trask B.J., Morris C.A.,  
RA Keating M.T.;  
RT "Complete physical map of the common deletion region in Williams  
RT syndrome and identification and characterization of three novel  
RT genes.";  
RL Hum. Genet. 103:590-599(1998).  
RN [4]  
RP SEQUENCE FROM N.A. (ISOFORM 6).  
RC TISSUE=Eye;  
RA Strausberg R.;  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: Transcriptional repressor. Binds to the canonical and  
CC non-canonical E box sequences 5'-CACGTG-3' (by similarity).  
CC -1- SUBUNIT: Binds DNA as a heterodimer with TCFL4/MLX.  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- ALTERNATIVE PRODUCTS: 6 ISOFORMS; 1/ALPHA (SHOWN HERE), 2/BETA,  
CC 3/GAMMA, 4/DELTA, 5/EPSILON AND 6; ARE PRODUCED BY ALTERNATIVE  
CC SPLICING.  
CC -1- TISSUE SPECIFICITY: EXPRESSED IN LIVER, HEART, KIDNEY, CEREBELLUM  
CC AND INTESTINAL TISSUES.  
CC -1- DISEASE: Haploinsufficiency of WBSCR14 may be the cause of certain  
CC cardiovascular and musculo-skeletal abnormalities observed in  
CC Williams-Beuren syndrome (WBS), a rare developmental disorder. It  
CC is a contiguous gene deletion syndrome involving genes from  
CC chromosome band 7q11.23.  
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
CC TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: AF156673; AAF68176.1; -  
CC EMBL: AF156603; AAF68174.1; -  
CC EMBL: AF245470; AAK20935.1; -  
CC EMBL: AF245471; AAK20936.1; -  
CC EMBL: AF245472; AAK20937.1; -  
CC EMBL: AF245473; AAK20938.1; -



DR EMBL; AF245474; AAK20939.1; -  
 DR EMBL; AF056184; AAD28084.1; -  
 DR EMBL; BC012925; AAH12925.1; -  
 DR HSSP; P25912; IHLO.  
 DR Genew; HGNC:12744; WBSR14.  
 DR MIN; 605678; -  
 DR MIN; 194050; -  
 DR InterPro: IPR001092; HLH\_basic.  
 DR Pfam; PF00010; HLH; 4.  
 DR SMART; SM00353; HLH; 1.  
 DR PROSITE; PS00038; HLH\_1; 1.  
 DR PROSITE; PS50888; HLH\_2; 1.  
 KW Transcription regulation; Repressor; Nuclear protein; DNA-binding;  
 KW Williams-Beuren syndrome; Alternative splicing.  
 FT DOMAIN 387 394 POLY-PRO.  
 FT DOMAIN 409 417 POLY-PRO.  
 FT DNA\_BIND 648 662 BASIC DOMAIN.  
 FT DOMAIN 688 702 HELIX-LOOP-HELIX MOTIF (POTENTIAL).  
 FT DOMAIN 703 724 LEUCINE-ZIPPER.  
 FT VARSPLIC 176 268 MISSING (IN ISOFORM 6).  
 FT VARSPLIC 558 575 QETVPEPCTFLPTTPAP -> AVNGCGQTSAPCQALGL  
 (IN ISOFORM 5).  
 FT VARSPLIC 576 852 MISSING (IN ISOFORM 5).  
 FT VARSPLIC 647 648 MISSING (IN ISOFORM 3 AND ISOFORM 4).  
 FT VARSPLIC 687 705 MISSING (IN ISOFORM 2 AND ISOFORM 4).  
 FT VARSPLIC 814 815 TV -> ST (IN ISOFORM 6).  
 FT VARSPLIC 816 852 MISSING (IN ISOFORM 6).  
 FT CONFLICT 558 558 MISSING (IN REF. 4).  
 SQ SEQUENCE 852 AA; 93071 MW; D495C3D7C0A72EC CRC64;  
 Query Match 68.4%; Score 39; DB 1; Length 852;  
 Best Local Similarity 83.3%; Pred. No. 1.2e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Qy 3 RWPPQ 8  
 Db 210 RWPPPE 215  
 RESULT 9  
 T13C\_HUMAN  
 ID T13C\_HUMAN STANDARD; PRT; 184 AA.  
 AC Q96R3;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Tumor necrosis factor receptor superfamily member 13C (B cell-  
 activating factor receptor) (BAFF receptor) (BAFF-R) (Blys receptor  
 3).  
 GN TNFRSF13C OR BAFFR OR BR3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 [1]  
 RN SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 RC TISSUE-B-cell lymphoma;  
 RX MEDLINE=21442025; PubMed=11509692;  
 RA Thompson J.S., Bixler S.A., Qian F., Vora K., Scott M.L.,  
 RA Cachero T.G., Hession C., Schneider P., Sizing I.D., Mullen C.,  
 RA Strauch K., Zafari M., Benjamin C.D., Tschopp J., Browning J.L.,  
 RA Ambrose C.;  
 RT "BAFF-R, a newly identified TNF receptor that specifically interacts  
 with BAFF";  
 RT Science 293:2108-2111(2001).  
 RN [2]  
 RN FUNCTION.  
 RP MEDLINE=21475520; PubMed=11591325;  
 RX Yan M., Brady J.R., Chan B., Lee W.P., Hsu B., Harless S.M.,  
 RA Cancro M.P., Grewal I.S., Dixit V.M.;  
 RT "Identification of a novel receptor for B lymphocyte stimulator that  
 is mutated in a mouse strain with severe B cell deficiency";  
 RL Curr. Biol. 11:1547-1552(2001).

CC -1- FUNCTION: B-cell receptor specific for TNFSF13B/TALL1/BAFF/BLYS.  
 CC Promotes the survival of mature B-cells and the B-cell response.  
 CC -1- SUBCELLULAR LOCATION: Type III membrane protein (Probable).  
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be  
 CC produced by alternative splicing.  
 CC -1- TISSUE SPECIFICITY: Highly expressed in spleen and lymph node, and  
 CC in resting B-cells. Detected at lower levels in activated B-cells,  
 CC resting CD4+ T-cells, in thymus and peripheral blood leukocytes.  
 CC -1- SIMILARITY: CONTAINS 1 TNFR-CYS REPEAT.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF373846; AAK91826.1; -  
 DR Genew; HGNC:17755; TNFRSF13C.  
 DR MIM; 606269; -  
 DR PROSITE; PS00652; TNFR\_NGFR\_1; FALSE\_NEG.  
 DR PROSITE; PS50050; TNFR\_NGFR\_2; FALSE\_NEG.  
 KW Receptor; Immune response; Signal-anchor; Transmembrane;  
 KW Alternative splicing.  
 FT DOMAIN 1 78 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 79 99 SIGNAL-ANCHOR (TYPE III MEMBRANE PROTEIN)  
 (POTENTIAL).  
 FT DOMAIN 100 184 CYTOPLASMIC (POTENTIAL).  
 FT REPEAT 18 35 TNFR-CYS (PARTIAL).  
 FT DISULFID 19 32 BY SIMILARITY.  
 FT DISULFID 24 35 BY SIMILARITY.  
 FT VARSPLIC 143 143 P -> PA (IN ISOFORM 2).  
 SQ SEQUENCE 184 AA; 18863 MW; F2BF98099A27138 CRC64;  
 Query Match 66.7%; Score 38; DB 1; Length 184;  
 Best Local Similarity 83.3%; Pred. No. 37;  
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 2 RWPPP 7  
 Db 143 PAWPPP 148  
 RESULT 10  
 Y497\_MYCTU  
 ID Y497\_MYCTU STANDARD; PRT; 310 AA.  
 AC Q11162;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical protein Rv0497  
 GN RV0497 OR MT0517 OR MTCY20G9.23.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
 OC Actinomycetales; Corynebacteriales; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Hoilroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence";  
 RL Nature 393:537-544(1998).  
 RN [2]

```

RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: TO M.LEPRAE ML2433.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: 277162; CAB00923.1; -
CC EMBL: AE006952; AAK44740.1; -
CC TIGR: MT0517; -
CC DR TubercuList; RV0497; -
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 231 251 POTENTIAL.
FT TRANSMEM 257 277 POTENTIAL.
FT TRANSMEM 286 306 POTENTIAL.
FT DOMAIN 33 39 POLY-ARG.
FT DOMAIN 197 202 POLY-ALA.
SQ SEQUENCE 310 AA; 33092 MW; 4954027F694DF5C2 CRC64;

Query Match 66.7%; Score 38; DB 1; Length 310;
Best Local Similarity 77.8%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

QY 2 PRWP--PPQ 8
DB 113 PRWPKSPQP 121

RESULT 11
AB13_ARATH
ID AB13_ARATH STANDARD; PRT; 720 AA.
AC Q01593;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Absciscic acid-insensitive protein 3.
GN AB13 OR AT3G24650 OR MSD24.2.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia; TISSUE=Fruit;
RX MEDLINE=93076109; PubMed=1359917;
RA Giraudat J., Hauge B.M., Valon C., Smalle J., Parcy F.,
RA Goodman H.M.;
RT "Isolation of the Arabidopsis AB13 gene by positional cloning.";
RL Plant Cell 4:1251-1261(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20363099; PubMed=10907853;
RA Kaneko T., Katoh T., Sato S., Nakamura A., Asamizu E., Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the 4,251,695 bp regions covered by 90 Pl, TAC
RT and BAC clones.";
RL DNA Res. 7:217-221(2000).

```

```

CC -!- FUNCTION: COULD PARTICIPATE IN ABSICISIC ACID-REGULATED GENE
CC EXPRESSION DURING SEED DEVELOPMENT.
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- SIMILARITY: CONTAINS 1 TF-B3 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X68141; CAA48241.1; -
CC EMBL: AF000740; BAB01214.1; -
CC PIR: J01676; J01676.
CC InterPro: IPR003340; TF_B3.
CC Pfam: PF02362; B3; 1.
CC Developmental protein; Transcription regulation; DNA-binding;
KW Activator; Nuclear protein.
FT DOMAIN 532 720 TF-B3.
SQ SEQUENCE 720 AA; 79500 MW; D6A0F740D733060F CRC64;

Query Match 66.7%; Score 38; DB 1; Length 720;
Best Local Similarity 83.3%; Pred. No. 1.5e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 PRWPPP 7
DB 391 PSMPPP 396

RESULT 12
WDRA_HUMAN
ID WDRA_HUMAN STANDARD; PRT; 1242 AA.
AC Q9HBG6;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE WD-repeat protein 10.
GN WDR10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21134653; PubMed=11242542;
RA Gross C., De Baere E., Lo A., Chang W., Messiaen L.;
RT "Cloning and characterization of human WDR10, a novel gene located at
RT 3q21 encoding a WD-repeat protein that is highly expressed in
RT pituitary and testis.";
RL DNA Cell Biol. 20:41-52(2001).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- TISSUE SPECIFICITY: Expressed in many tissues. Predominant
CC expression in testis and pituitary.
CC -!- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF244930; AAG15427.1; -
CC Genbank: HGNC:13556; WDR10.
CC MIM: 605045; -
CC InterPro: IPR001680; WD40.
CC Pfam: PF00400; WD40; 7.
CC ProDom: PD000018; WD40; 1.
CC PROSITE: PS00678; WD_REPEATS_1; FALSE_NEG.

```



ions, magnesium ions, ATP and calmodulin.  
 CC -!- MISCELLANEOUS: The calcium release channel activity resides in the  
 CC C-terminal region while the remaining part of the protein  
 CC constitutes the 'foot' structure spanning the junctional gap  
 CC between the SR and the T-tubule. It is possible that the foot  
 CC structure interacts with the cytoplasmic region of the  
 CC dihydropyridine receptor.  
 CC -!- MISCELLANEOUS: Ryanodine is an alkaloid that binds to the Ca-  
 CC release channel in junctional SR and modulates its activity.  
 CC -!- SIMILARITY: BELONGS TO THE RYANODINE RECEPTOR FAMILY.  
 CC -!- SIMILARITY: CONTAINS 3 SPRY DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X62880; CAA44674.1; ALT\_SEQ.  
 DR EMBL; X68247; CAA48318.1; -  
 DR EMBL; X69465; CAA49225.1; -  
 DR EMBL; M32501; AAA31022.1; -  
 DR PIR; A37105; A37105.  
 DR InterPro; IPR000699; Ca-rel\_channel.  
 DR InterPro; IPR01682; Ca/Na\_pore.  
 DR InterPro; IPR002048; EF-hand.  
 DR InterPro; IPR000636; M-channel\_nlg.  
 DR InterPro; IPR003608; MIR.  
 DR InterPro; IPR003032; Ryr.  
 DR InterPro; IPR001215; Ryanodn\_receptor.  
 DR InterPro; IPR003878; SPRY\_domain.  
 DR InterPro; IPR003877; SPRY\_receptor.  
 DR Pfam; PF00520; ion\_trans; 1.  
 DR Pfam; PF00622; SPRY; 3.  
 DR Pfam; PF01365; RYDR\_ITPR; 2.  
 DR Pfam; PF02026; Ryr; 4.  
 DR Pfam; PF02815; MIR; 4.  
 DR PRINTS; PR00795; RYANODINER.  
 DR SMART; SM00472; MIR; 4.  
 DR SMART; SM00449; SPRY; 3.  
 KW Receptor; Transmembrane; Ionic channel; Calcium channel; Repeat;  
 KW Phosphorylation; Glycoprotein.  
 FT DOMAIN 1 3124  
 FT TRANSMEM 3125 3145 M' (POTENTIAL).  
 FT TRANSMEM 3189 3207 M' (POTENTIAL).  
 FT TRANSMEM 3980 3999 M1 (POTENTIAL).  
 FT TRANSMEM 4018 4036 M2 (POTENTIAL).  
 FT TRANSMEM 4274 4297 M3 (POTENTIAL).  
 FT TRANSMEM 4339 4359 M4 (POTENTIAL).  
 FT TRANSMEM 4557 4578 M5 (POTENTIAL).  
 FT TRANSMEM 4646 4669 M6 (POTENTIAL).  
 FT TRANSMEM 4767 4807 M7 (POTENTIAL).  
 FT TRANSMEM 4835 4854 M8 (POTENTIAL).  
 FT TRANSMEM 4877 4896 M9 (POTENTIAL).  
 FT TRANSMEM 4921 4935 M10 (POTENTIAL).  
 FT DOMAIN 659 797 SPRY 1.  
 FT DOMAIN 1085 1208 SPRY 2.  
 FT DOMAIN 1430 1570 SPRY 3.  
 FT DOMAIN 1874 1925 GLU-RICH (ACIDIC).  
 FT DOMAIN 4458 4526 PRO-RICH.  
 FT DOMAIN 842 2960 6 X APPROXIMATE REPEATS.  
 FT REPEAT 842 955 1.  
 FT REPEAT 956 1069 2.  
 FT REPEAT 1345 1360 3 (INCOMPLETE).  
 FT REPEAT 1373 1388 4 (INCOMPLETE).  
 FT REPEAT 2727 2846 5.  
 FT REPEAT 2847 2960 6.  
 FT MOD\_RES 2844 2844 PHOSPHORYLATION (BY PKA AND PKG) (BY  
 FT MOD\_RES 3947 3947 SIMILARITY).  
 FT MOD\_RES 4320 4320 PHOSPHORYLATION (POTENTIAL).  
 FT MOD\_RES 4320 4320 PHOSPHORYLATION (POTENTIAL).

FT CARBOHYD 3467 3467 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 3475 3475 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 3904 3904 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 3945 3945 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 4144 4144 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 4862 4862 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 2092 A -> P (IN REF. 2 AND 3).  
 SQ SEQUENCE 5035 AA; 565317 MW; E00613F2027B94A4 CRC64;  
 Query Match 65.8%; Score 37.5; DB 1; Length 5035;  
 Best Local Similarity 50.0%; Pred. No. 1.2e+03;  
 Matches 8; Conservative 0; Mismatches 1; Indels 7; Gaps 1;  
 Qy 1 LPRW-----PPQL 9  
 DB 3282 LPRWGEPEAPPAL 3297  
 RESULT 15  
 RYR1\_RABIT STANDARD; PRT; 5037 AA.  
 ID RYR1\_RABIT  
 AC P11716;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Ryanodine receptor 1 (Skeletal muscle-type ryanodine receptor) (RyR1)  
 DE (RYR-1) (Skeletal muscle calcium release channel).  
 GN RYR1.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE-Skeletal muscle;  
 RX MEDLINE=89262082; PubMed=2725677;  
 RA Takeshima H., Nishimura S., Matsumoto T., Ishido H., Kangawa K.,  
 RA Minamino N., Matsuo H., Ueda M., Hanaka M., Hirose T., Numa S.;  
 RT "Primary structure and expression from complementary DNA of skeletal  
 RT muscle ryanodine receptor.";  
 RL Nature 339:439-445(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Skeletal muscle;  
 RX MEDLINE=90130482; PubMed=2298749;  
 RA Zorzato F., Fujii J., Otsu K., Phillips M., Green N.M., Lai P.A.,  
 RA Meissner G., MacLennan D.H.;  
 RT "Molecular cloning of cDNA encoding human and rabbit forms of the  
 RT Ca2+ release channel (ryanodine receptor) of skeletal muscle  
 RT sarcoplasmic reticulum.";  
 RL J. Biol. Chem. 265:2244-2256(1990).  
 RN [3]  
 RP PHOSPHORYLATION OF SER-2843.  
 RX MEDLINE=93120178; PubMed=8380342;  
 RA Suko J., Maurer-Fogy I., Plank B., Bertel O., Wyskovsky W.,  
 RA Hohenegger M., Hellmann G.;  
 RT "Phosphorylation of serine 2843 in ryanodine receptor-calcium release  
 RT channel of skeletal muscle by cAMP-, cGMP- and CaM-dependent protein  
 RT kinase.";  
 RL Biochim. Biophys. Acta 1175:193-206(1993).  
 CC -!- FUNCTION: COMMUNICATION BETWEEN TRANSVERSE-TUBULES AND  
 CC SARCOPLASMIC RETICULUM. CONTRACTION OF SKELETAL MUSCLE IS  
 CC TRIGGERED BY RELEASE OF CALCIUM IONS FROM SR FOLLOWING  
 CC DEPOLARIZATION OF T-TUBULES.  
 CC -!- SUBUNIT: HOMOTETRAMER (POTENTIAL).  
 CC -!- TISSUE SPECIFICITY: FAST- OR SLOW-TWITCH SKELETAL MUSCLE.  
 CC -!- PTM: The N-terminus is blocked.  
 CC -!- MISCELLANEOUS: The calcium release channel is modulated by calcium  
 CC ions, magnesium ions, ATP and calmodulin.  
 CC -!- MISCELLANEOUS: The calcium release channel activity resides in the  
 CC C-terminal region while the remaining part of the protein  
 CC constitutes the 'foot' structure spanning the junctional gap  
 CC between the SR and the T-tubule. It is possible that the foot

CC structure interacts with the cytoplasmic region of the  
CC dihydropyridine receptor.  
CC -1- MISCELLANEOUS: Ryanodine is an alkaloid that binds to the Ca-  
CC release channel in junctional SR and modulates its activity.  
CC -1- SIMILARITY: BELONGS TO THE RYANODINE RECEPTOR FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: X15209; CAA33279.1; -;  
CC EMBL: X15750; CAA33762.1; -;  
CC PIR: S04654; S04654.  
CC InterPro: IPR000699; Ca-rel\_channel.  
DR InterPro: IPR001682; Ca/Na\_pore.  
DR InterPro: IPR002048; EF-hand.  
DR InterPro: IPR000636; M-channel\_nlg.  
DR InterPro: IPR003608; MIR.  
DR InterPro: IPR003032; Ryr.  
DR InterPro: IPR001215; Ryanodn\_receptor.  
DR InterPro: IPR003878; SPRY\_domain.  
DR InterPro: IPR003877; SPRY\_receptor.  
DR Pfam: PF00520; Ion\_trans; 1.  
DR Pfam: PF00622; SPRY; 3.  
DR Pfam: PF01365; RYDR\_ITPR; 2.  
DR Pfam: PF02026; RYR; 4.  
DR Pfam: PF02815; MIR; 4.  
DR PRINTS: PR00795; RYANODINER.  
DR SMART: SM00472; MIR; 4.  
DR SMART: SM00449; SPRY; 3.  
KW Receptor; Transmembrane; Ionic channel; Calcium channel; Repeat;  
KW Phosphorylation; Glycoprotein.  
FT DOMAIN 1 3123  
FT TRANSMEM 3124 3144  
FT M' (POTENTIAL).  
FT TRANSMEM 3188 3206  
FT M'' (POTENTIAL).  
FT TRANSMEM 3985 4004  
FT M1 (POTENTIAL).  
FT TRANSMEM 4023 4041  
FT M2 (POTENTIAL).  
FT TRANSMEM 4277 4300  
FT M3 (POTENTIAL).  
FT TRANSMEM 4342 4362  
FT M4 (POTENTIAL).  
FT TRANSMEM 4559 4580  
FT M5 (POTENTIAL).  
FT TRANSMEM 4648 4671  
FT M6 (POTENTIAL).  
FT TRANSMEM 4789 4809  
FT M7 (POTENTIAL).  
FT TRANSMEM 4837 4856  
FT M8 (POTENTIAL).  
FT TRANSMEM 4879 4898  
FT M9 (POTENTIAL).  
FT TRANSMEM 4914 4937  
FT M10 (POTENTIAL).  
FT SPRY 1.  
FT DOMAIN 659 797  
FT SPRY 2.  
FT DOMAIN 1085 1208  
FT SPRY 3.  
FT DOMAIN 1430 1570  
FT GLU-RICH (ACIDIC).  
FT DOMAIN 1873 1913  
FT REPEAT 842 2959  
FT 6 X APPROXIMATE REPEATS.  
FT REPEAT 842 955  
FT 1.  
FT REPEAT 956 1069  
FT 2.  
FT REPEAT 1345 1360  
FT 3 (INCOMPLETE).  
FT REPEAT 1373 1388  
FT 4 (INCOMPLETE).  
FT REPEAT 2726 2845  
FT 5.  
FT REPEAT 2846 2959  
FT 6.  
FT MOD\_RES 2843 2843  
FT PHOSPHORYLATION (BY PKA AND PKG).  
FT MOD\_RES 3952 3952  
FT PHOSPHORYLATION (POTENTIAL).  
FT MOD\_RES 4323 4323  
FT WITH ACHR-SUBUNITS TRANSMEMBRANE SEGMENTS  
FT SIMILAR 4628 4861  
FT M2 AND M3.  
FT CARBOHYD 3466  
FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 3909  
FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 3950  
FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 4149  
FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 4864  
FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CONFLICT 2015 2015  
FT E -> D (IN REF. 2).  
FT CONFLICT 3481 3485  
FT MISSING (IN REF. 2).  
SQ SEQUENCE 5037 AA; 565238 MW; 4ABD87AA26697070 CRC64;

Query Match 65.8%; Score 37.5; DB 1; Length 5037;  
Best Local Similarity 50.0%; Pred. No. 1.2e+03;  
Matches 8; Conservative 0; Mismatches 1; Indels 7; Gaps 1;  
Qy 1 LPRW-----PPQQL 9  
Db 3281 LPRWWRGPEAPPAL 3296  
Search completed: November 26, 2002, 12:02:26  
Job time : 7.90323 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 26, 2002, 11:58:11 ; Search time 4.93548 Seconds  
(without alignments)  
175.304 Million cell updates/sec

Title: US-09-674-593-3  
Perfect score: 57  
Sequence: 1 LPRWPPPPQL 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	77.2	1733	1 B45344	probable nuclear a
2	44	77.2	1958	2 B40505	hypothetical prote
3	42	73.7	211	2 C84681	hypothetical prote
4	41	71.9	190	2 F75336	conserved hypotet
5	41	71.9	211	2 T49803	hypothetical prote
6	41	71.9	410	2 B35670	protein-tyrosine k
7	41	71.9	789	1 A45617	6-phosphofructokin
8	40	70.2	61	2 T03587	isocitrate dehydro
9	40	70.2	166	2 E90819	hypothetical prote
10	40	70.2	180	2 T25460	hypothetical prote
11	40	70.2	206	2 A87253	hypothetical prote
12	40	70.2	218	2 G72754	probable phosphoen
13	40	70.2	409	2 T24543	hypothetical prote
14	40	70.2	578	2 T22888	hypothetical prote
15	40	70.2	592	2 T22885	hypothetical prote
16	40	70.2	629	2 T22890	hypothetical prote
17	40	70.2	1274	2 T37193	enamelin matrix pr
18	39	68.4	197	2 A70750	hypothetical prote
19	39	68.4	280	2 T17116	protein kinase cdc
20	39	68.4	299	2 AF3045	transcription regu
21	39	68.4	299	2 F98240	gstR protein (AF0)
22	39	68.4	613	2 JC7762	SOX-3 protein (AF0
23	39	68.4	669	2 T51246	ARR1 protein [limp
24	38	66.7	104	2 S53993	hypothetical prote
25	38	66.7	153	2 G72481	probable histidine
26	38	66.7	304	2 B69304	3-hydroxyacyl-CoA
27	38	66.7	310	2 D70745	hypothetical prote
28	38	66.7	720	2 J01676	ABI3 protein - Ara
29	38	66.7	1198	2 T43484	hypothetical prote

30	38	66.7	2411	2 A46299	tyrosine kinase su
31	37.5	65.8	203	2 AG3488	protein ybis precu
32	37.5	65.8	1520	2 T00273	hypothetical prote
33	37.5	65.8	5035	1 I46646	ryanodine receptor
34	37.5	65.8	5037	2 B35041	ryanodine receptor
35	37	64.9	13	2 A05174	tryptophyllin-13 -
36	37	64.9	134	2 D84672	hypothetical prote
37	37	64.9	142	2 C72523	hypothetical prote
38	37	64.9	211	2 A46458	human CRI homolog
39	37	64.9	218	2 A84688	hypothetical prote
40	37	64.9	251	2 T26856	hypothetical prote
41	37	64.9	264	2 E72642	hypothetical prote
42	37	64.9	273	2 AF2922	conserved hypotet
43	37	64.9	287	2 G97696	hypothetical prote
44	37	64.9	288	2 T06282	alpha-gliadin prec
45	37	64.9	294	1 J02243	protein kinase (EC

ALIGNMENTS

RESULT 1

B45344  
Probable nuclear antigen - suid herpesvirus 1 (strain Kaplan)  
C:Species: suid herpesvirus 1  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 16-Jul-1999  
C:Accession: B45344  
R:Vicek, C.; Kozmik, Z.; Paces, V.; Schirm, S.; Schwytzer, M.  
Virology 179, 365-377, 1990  
A:Title: Pseudorabies virus immediate-early gene overlaps with an oppositely oriented  
A:Reference number: A45344; MUID:91021039; PMID:2171211  
A:Accession: B45344  
A>Status: translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1733 <VLC>  
A:Cross-references: GB:M34651; NID:g334070; PIDN:AAA47471.1; PID:g334072  
C:Superfamily: pseudorabies virus 1 nuclear antigen

Query Match 77.2%; Score 44; DB 1; Length 1733;  
Best Local Similarity 100.0%; Pred. No. 87;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 PRWPPP 7  
| | | | |  
Db 213 PRWPPP 218

RESULT 2

B40505  
hypothetical protein - suid herpesvirus 1 (strain Indiana-Funkhuser or Becker)  
C:Species: suid herpesvirus 1  
C:Date: 10-Apr-1992 #sequence\_revision 10-Apr-1992 #text\_change 02-Sep-2000  
C:Accession: B40505  
R:Cheung, A.K.  
J. Virol. 65, 5260-5271, 1991  
A:Title: Cloning of the latency gene and the early protein 0 gene of pseudorabies vir  
A:Reference number: A40505; MUID:91374576; PMID:1654441  
A:Accession: B40505  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1958 <CHE>  
A:Cross-references: GB:M57505; NID:g334066; PIDN:AAA47468.1; PID:g334068  
C:Superfamily: pseudorabies virus 1 nuclear antigen

Query Match 77.2%; Score 44; DB 2; Length 1958;  
Best Local Similarity 100.0%; Pred. No. 99;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 PRWPPP 7  
| | | | |  
Db 423 PRWPPP 428

## RESULT 3

C84691  
 A:Cross-references: EMBL:AL356834; GSPDB:GN00116; NCSP:BL1B22.50  
 A:Experimental source: BAC clone B1B22; strain OR74A  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
 C:Accession: C84681  
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon, L.;  
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
 Nature 402, 761-768, 1999  
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A:Reference number: A84420; MUID:20083487; PMID:10617197  
 A:Accession: C84681  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-211 <SFO>  
 A:Cross-references: GB:AE002093; NID:g4063744; PIDN:AAC98452.1; GSPDB:GN00139  
 C:Genetics:  
 A:Gene: At2g28140  
 A:Map position: 2

Query Match 73.7%; Score 42; DB 2; Length 211;  
 Best Local Similarity 72.7%; Pred. No. 19;  
 Matches 8; Conservative 1; Mismatches 0; Indels 2; Gaps 1;

QY 1 LPRW-PPQOL 9  
 ||| |||||  
 Db 156 LPQWRHPPQOL 166

## RESULT 4

F75336.  
 A:Cross-references: GB:AE002032; NID:g6459715; PIDN:AAF11472.1; PID:g645970  
 A:Experimental source: strain R1  
 C:Genetics:  
 A:Gene: DR1919  
 A:Map position: 1

Query Match 71.9%; Score 41; DB 2; Length 190;  
 Best Local Similarity 83.3%; Pred. No. 25;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 PRWPPP 7  
 |||||  
 Db 17 PKWPPP 22

## RESULT 5

T49803  
 A:Cross-references: GB:AE002032; NID:g6459715; PIDN:AAF11472.1; PID:g645970  
 A:Experimental source: strain R1  
 C:Genetics:  
 A:Gene: DR1919  
 A:Map position: 1

Query Match 71.9%; Score 41; DB 1; Length 789;  
 Best Local Similarity 71.4%; Pred. No. 1.1e-02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

A:Residues: 1-211 <SCH>

A:Cross-references: EMBL:AL356834; GSPDB:GN00116; NCSP:BL1B22.50  
 A:Experimental source: BAC clone B1B22; strain OR74A  
 C:Genetics:

A:Gene: NCSP:BL1B22.50

A:Map position: 6

C:Superfamily: Neurospora crassa hypothetical protein B1B22.50

Query Match 71.9%; Score 41; DB 2; Length 211;  
 Best Local Similarity 100.0%; Pred. No. 27;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPRWPP 6  
 |||||  
 Db 147 LPRWPP 152

## RESULT 6

B35670  
 A:Cross-references: EC 2.7.1.112) 2 - slime mold (Dictyostelium discoideum) (fra  
 C:Species: Dictyostelium discoideum  
 C>Date: 28-Sep-1990 #sequence\_revision 28-Sep-1990 #text\_change 04-Feb-2000  
 C:Accession: B35670  
 R:Tan, J.L.; Spudich, J.A.  
 Mol. Cell. Biol. 10, 3578-3583, 1990  
 A:Title: Developmentally regulated protein-tyrosine kinase genes in Dictyostelium dis  
 A:Reference number: A35670; MUID:90287147; PMID:1972546  
 A:Accession: B35670  
 A>Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-410 <TAN>

A:Cross-references: GB:M33784; NID:g167777; PIDN:AAA3203.1; PID:g167778  
 C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom  
 C:Keywords: ATP; phosphotransferase; tyrosine-specific protein kinase  
 F:106-371/Domain: protein kinase homology <KIN>  
 F:114-122/Region: protein kinase ATP-binding motif

Query Match 71.9%; Score 41; DB 2; Length 410;  
 Best Local Similarity 75.0%; Pred. No. 55;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 PRWPPQOL 9  
 ||| |||  
 Db 273 PRWRPPEL 280

## RESULT 7

A45617  
 A:Cross-references: EC 2.7.1.11) - nematode (Haemonchus contortus)  
 N:Alternate names: phosphofructokinase 1; phosphohexokinase  
 C:Species: Haemonchus contortus  
 C>Date: 03-Feb-1994 #sequence\_revision 14-Jul-1994 #text\_change 11-Jun-1999  
 C:Accession: A45617  
 R:Klein, R.D.; Olson, E.R.; Favreau, M.A.; Winterrowd, C.A.; Hatzenbuehler, N.T.; Shea  
 Mol. Biochem. Parasitol. 48, 17-26, 1991  
 A:Title: Cloning of a cDNA encoding phosphofructokinase from Haemonchus contortus.  
 A:Reference number: A45617; MUID:92140452; PMID:1838137  
 A:Accession: A45617  
 A:Molecule type: mRNA  
 A:Residues: 1-789 <KLE>

A:Cross-references: GB:M59805; NID:g159184; PIDN:AAA29181.1; PID:g159185  
 A:Experimental source: strain DF1020  
 A>Note: this sequence is inconsistent with the nucleotide translation  
 C:Superfamily: human 6-phosphofructokinase; 6-phosphofructokinase 1 homology  
 C:Keywords: ATP; glycolysis; phosphoprotein; phosphotransferase  
 F:34-342/Domain: 6-phosphofructokinase 1 homology <6PF1>  
 F:421-708/Domain: 6-phosphofructokinase 1 homology <6PF2>

Query Match 71.9%; Score 41; DB 1; Length 789;  
 Best Local Similarity 71.4%; Pred. No. 1.1e-02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;



Qy 1 LPRWPPP 7  
:|||||  
Db 246 IPWPPPP 252

RESULT 8  
T03587  
isocitrate dehydrogenase homolog - rice (fragment)  
C:Species: Oryza sativa (rice)  
C>Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 21-Jul-2000  
C:Accession: T03587  
R:Umeda, M.; Hara, C.; Matsubayashi, Y.; Li, H.; Lui, Q.; Tadokoro, F.; Aotsuka, S.; Uchi  
Plant Mol. Biol. 25, 469-478, 1994  
A>Title: Expressed sequence tags from cultured cells of rice (*Oryza sativa* L.) under str  
A:Reference number: 214962; MUID:94325471; PMID:8049371  
A:Accession: T03587  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-61 <OMES>  
A:Cross-references: EMBL:D21069; NID:9413915; PIDN:BAA04645.1; PID:9602777  
A:Experimental source: callus  
C:Genetics:  
A:Note: AK88  
C:Superfamily: yeast isocitrate dehydrogenase (NADPP+)

Query Match 70.2%; Score 40; DB 2; Length 61;  
Best Local Similarity 75.0%; Pred. No. 11;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 PRWPPPPQ 9  
:|||||  
Db 15 PAWPPPPWL 22

RESULT 9  
E90819  
hypothetical protein ECS1526 [imported] - Escherichia coli (strain O157:H7, substrain RI  
C:Species: Escherichia coli  
C>Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
C:Accession: E90819  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc  
A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: E90819  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-166 <HAY>  
A:Cross-references: GB:BA000007; PIDN:BA034948.1; PID:q13360989; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain RIMD 0509952  
C:Genetics:  
A:Gene: ECS1526

Query Match 70.2%; Score 40; DB 2; Length 166;  
Best Local Similarity 83.3%; Pred. No. 30;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 PRWPPPP 7  
:|||||  
Db 135 POWPPPP 140

RESULT 10  
T25460  
hypothetical protein B0432.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jun-2000  
C:Accession: T25460  
R:Henkhaus, J.; Wohldmann, P.  
Submitted to the EMBL Data Library, December 1996  
A:Description: The sequence of C. elegans cosmid B0432.  
A:Reference number: Z20038

A:Accession: T25460  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-180 <HEN>  
A:Cross-references: EMBL:U08036; PIDN:AAB37899.1; GSPDB:GN00020; CESP:B0432.3  
A:Experimental source: strain Bristol N2; clone B0432  
C:Genetics:  
A:Gene: CESP:B0432.3  
A:Map position: 2  
A:Introns: 70/2; 106/2; 148/2  
C:Superfamily: Caenorhabditis elegans hypothetical protein B0432.3

Query Match 70.2%; Score 40; DB 2; Length 180;  
Best Local Similarity 71.4%; Pred. No. 33;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 3 RWPPPPQ 9  
:|||||  
Db 148 RWPPPPKM 154

RESULT 11  
A87253  
hypothetical protein CC0033 [imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C>Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
C:Accession: A87253  
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, K.  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; K  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A>Title: Complete Genome Sequence of Caulobacter crescentus.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: A87253  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-206 <STO>  
A:Cross-references: GB:AB005673; NID:q13421125; PIDN:AAK22021.1; GSPDB:GN00148  
C:Genetics:  
A:Gene: CC0033

Query Match 70.2%; Score 40; DB 2; Length 206;  
Best Local Similarity 85.7%; Pred. No. 38;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LPRWPPP 7  
:|||||  
Db 126 LPRWPEP 132

RESULT 12  
G72754  
probable phosphoenolpyruvate synthase APE0028 - Aeropyrum pernix (strain K1)  
C:Species: Aeropyrum pernix  
C>Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
C:Accession: G72754  
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Ta  
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.  
DNA Res. 6, 83-101, 1999  
A>Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero  
A:Reference number: A72450; MUID:99310339; PMID:10382966  
A:Accession: G72754  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-218 <RAW>  
A:Cross-references: DDBJ:AP000058; NID:g5103388; PIDN:BAA78937.1; PID:dl042713; PID:  
A:Experimental source: strain K1  
C:Genetics:  
A:Gene: APE0028

Query Match 70.2%; Score 40; DB 2; Length 218;  
Best Local Similarity 62.5%; Pred. No. 40;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LPRWPPQ 8  
:|||||  
Db 64 IQRWPPPE 71

## RESULT 13

T24543

hypothetical protein T05F1.8 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 02-Jun-2000

C:Accession: T24543

R:Burton, J.

submitted to the EMBL Data Library, November 1996

A:Reference number: Z19905

A:Accession: T24543

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-409 &lt;WIL&gt;

A:Cross-references: EMBL:Z81586; PIDN:CAB04697.1; GSPDB:GN00019; CESP:T05F1.8

A:Experimental source: clone T05f1

C:Genetics:

A:Gene: CESP:T05F1.8

A:Map position: 1

A:Introns: 337/3

C:Superfamily: ADP,ATP carrier protein; ADP,ATP carrier protein repeat homology

Query Match 70.2%; Score 40; DB 2; Length 409;

Best Local Similarity 77.8%; Pred. No. 77;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LPRWPPQ 9  
||| |||||  
Db 342 LPRPPPPQ 350

## RESULT 14

T22888

hypothetical protein F58A3.1b - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000

C:Accession: T22888

R:Lennard, N.

submitted to the EMBL Data Library, October 1996

A:Reference number: Z19632

A:Accession: T22888

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-578 &lt;WIL&gt;

A:Cross-references: EMBL:Z81017; PIDN:CAB02672.1; GSPDB:GN00028; CESP:F58A3.1b

A:Experimental source: clone F58A3

C:Genetics:

A:Gene: CESP:F58A3.1b

A:Map position: X

A:Introns: 30/2; 66/3; 108/1; 209/3; 256/3; 285/1; 356/3

Query Match 70.2%; Score 40; DB 2; Length 578;

Best Local Similarity 83.3%; Pred. No. 11e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 PRWPPP 7  
:|||||  
Db 566 POWPPP 571

## RESULT 15

T22885

hypothetical protein F58A3.1a - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000

C:Accession: T22885

R:Lennard, N.

submitted to the EMBL Data Library, October 1996

A:Reference number: Z19632

A:Accession: T22885

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-592 &lt;WIL&gt;

A:Cross-references: EMBL:Z81017; PIDN:CAB02669.1; GSPDB:GN00028; CESP:F58A3.1a

A:Experimental source: clone F58A3

C:Genetics:

A:Gene: CESP:F58A3.1a

A:Map position: X

A:Introns: 44/2; 80/3; 122/1; 223/3; 270/3; 299/1; 370/3

Query Match 70.2%; Score 40; DB 2; Length 592;

Best Local Similarity 83.3%; Pred. No. 1.1e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 PRWPPP 7  
:|||||  
Db 580 POWPPP 585

Search completed: November 26, 2002, 19:42:36

Job time : 10.9355 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2002, 17:06:07 ; Search time 9.48387 Seconds  
(without alignments)  
126.452 Million cell updates/sec

Title: US-09-674-593-3  
Perfect score: 57  
Sequence: 1 LPRWPPQL 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues  
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_101002.\*  
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	57	100.0	9	21	AA53810
2	57	100.0	84	21	AA53809
3	45	78.9	104	22	AA03578
4	45	78.9	195	22	AAU30711
5	44	77.2	90	22	ABG08009
6	44	77.2	443	22	AAU29704
7	44	77.2	709	19	AA58988
8	44	77.2	709	22	AA590668
9	44	77.2	1507	23	AB897356
10	44	77.2	1958	15	AA606020

11	43	75.4	53	23	ABP09141	Human ORFX protein
12	43	75.4	75	22	AA007833	Human polypeptide
13	42-5	74.6	32	22	ABG19493	Novel human diagno
14	42	73.7	142	22	ABG02119	Novel human diagno
15	42	73.7	429	22	ABG19033	Novel human diagno
16	42	73.7	2424	22	AB58924	Drosophila melanog
17	41	71.9	73	23	AA021709	Human secreted pro
18	41	71.9	204	22	ABG27363	Novel human diagno
19	41	71.9	223	21	AA841572	Human ORFX ORF1336
20	41	71.9	224	21	AA951810	Gene 28 human secr
21	41	71.9	228	22	AAU51500	Propionibacterium
22	41	71.9	789	12	AA15143	Haemophilus contort
23	41	71.9	978	22	AA65604	Novel protein kina
24	40	70.2	27	22	ABG15601	Novel human diagno
25	40	70.2	27	22	ABG15610	Novel human diagno
26	40	70.2	34	22	AA013538	Human polypeptide
27	40	70.2	46	21	AB32453	Human secreted pro
28	40	70.2	65	23	ABP08789	Human ORFX protein
29	40	70.2	79	22	ABG10176	Novel human diagno
30	40	70.2	108	22	AAU41561	Propionibacterium
31	40	70.2	285	22	AAW23514	Human EST encoded
32	40	70.2	336	21	AA827989	Human secreted pro
33	40	70.2	336	21	AA827990	Human secreted pro
34	40	70.2	336	22	AA64535	Gene 41 human secr
35	40	70.2	336	22	AA64536	Human secreted pro
36	40	70.2	398	22	ABG10177	Novel human diagno
37	40	70.2	457	22	AB663679	Drosophila melanog
38	40	70.2	514	22	ABG10181	Novel human diagno
39	40	70.2	571	22	AB67288	Drosophila melanog
40	40	70.2	585	20	AA13375	Amino acid sequenc
41	40	70.2	654	22	AAU29036	Human PRO polypept
42	40	70.2	654	22	AAE06593	Human protein havi
43	40	70.2	654	22	AA80243	Human PRO262 prote
44	40	70.2	796	22	AA894811	Human protein sequ
45	39	68.4	43	22	AA661771	Human brain expres

ALIGNMENTS

RESULT 1  
AA53810  
ID AA53810 standard; Peptide: 9 AA.  
XX  
AC AA53810;  
XX  
DT 22-FEB-2000 (first entry)  
XX  
DE Antigen encoded by tmour rejection antigen RUR-1 antisense cDNA.  
XX  
KW Antisense; human; tumour rejection antigen; RUR-1; tumour; cancer;  
KW renal cell carcinoma; colorectal carcinoma; melanoma; sarcoma;  
KW leukaemia.  
XX  
OS Homo sapiens.  
XX  
PN WO9958546-A1.  
XX  
PD 18-NOV-1999.  
XX  
PF 13-MAY-1999; 99WO-US10424.  
XX  
PR 13-MAY-1998; 98US-0085318.  
XX  
PA (LUDW-) LUDWIG INST CANCER RES.  
XX  
PI Van Den Eynde B, Boon-Falleur T;  
XX  
DR WPI: 2000-053076/04.  
XX  
PT New isolated tumour rejection antigen RUR-1 nucleic acids, used for,  
PT e.g. treatment of cancers -  
XX

PS Claim 17; Page 42; 75pp; English.

XX The present sequence represents an antigen derived from a protein  
CC encoded by the antisense cDNA sequence of human tumour rejection  
CC antigen RUR-1. The present antigen is recognised by CTL 361A/21.  
CC The cDNA sequence is the antisense strand of a ubiquitously expressed  
CC gene. The antisense strand codes for a polypeptide which is  
CC preferentially expressed in tumour samples and tumour-derived cells  
CC lines. The polypeptide is unrelated to any TRAP protein. The RUR-1  
CC antisense cDNA sequence was isolated from a renal cell carcinoma line  
CC LB9211-RCC. The RUR-1 nucleic acids and polypeptides can be used for  
CC diagnosis, prognosis or treatment of a disorder characterized by the  
CC expression of a RUR-1 antisense cDNA molecule or an expression product,  
CC such as cancers, e.g. renal cell carcinoma, colorectal carcinoma,  
CC melanoma, sarcoma or leukaemia.

XX SQ Sequence 9 AA;

Query Match 100.0%; Score 57; DB 21; Length 9;  
Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPRWPPQL 9  
| | | | | | | | |  
DB 1 LPRWPPQL 9

RESULT 2

AA53809  
ID AAY53809 standard; Protein; 84 AA.

AC AAY53809;

XX 22-FEB-2000 (first entry)

XX Protein encoded by tumour rejection antigen RUR-1 antisense cDNA.

XX Antisense; human; tumour rejection antigen; RUR-1; tumour; cancer;  
KW renal cell carcinoma; colorectal carcinoma; melanoma; sarcoma;  
KW leukaemia.

XX Homo sapiens.

XX WO9958546-A1.

XX 18-NOV-1999.

XX 13-MAY-1999; 99WO-US10424.

XX 13-MAY-1998; 98US-0085318.

XX (LUDW-) LUDWIG INST CANCER RES.

XX Van Den Eynde B, Boon-Falleur T;

XX WPI; 2000-053076/04.

XX N-PSDB; AAZ36643.

XX New isolated tumour rejection antigen RUR-1 nucleic acids, used for,  
PT e.g. treatment of cancers -

XX Claim 16; Fig 5; 75pp; English.

XX The present sequence is encoded by the antisense cDNA sequence of human  
CC tumour rejection antigen RUR-1. This cDNA sequence is the antisense  
CC strand of a ubiquitously expressed gene. The antisense strand codes for  
CC a polypeptide which is preferentially expressed in tumour samples and  
CC tumour-derived cells lines. The polypeptide is unrelated to any TRAP  
CC protein. The cDNA sequence was isolated from a renal cell carcinoma line  
CC LB9211-RCC. The RUR-1 nucleic acids and polypeptides can be used for  
CC diagnosis, prognosis or treatment of a disorder characterized by the  
CC expression of a RUR-1 antisense cDNA molecule or an expression product,  
CC such as cancers, e.g. renal cell carcinoma, colorectal carcinoma,

CC melanoma, sarcoma or leukaemia.

XX SQ Sequence 84 AA;

Query Match 100.0%; Score 57; DB 21; Length 84;  
Best Local Similarity 100.0%; Pred. No. 0.43;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPRWPPQL 9

| | | | | | | | |  
DB 38 LPRWPPQL 46

RESULT 3

AA03578  
ID AAO03578 standard; Protein; 104 AA.

XX AAO03578;

XX 06-NOV-2001 (first entry)

XX Human polypeptide SEQ ID NO 17470.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation.

XX Homo sapiens.

XX WO200164835-A2.

XX 07-SEP-2001.

XX 26-FEB-2001; 2001WO-US04927.

XX 28-FEB-2000; 2000US-0515126.

XX 18-MAY-2000; 2000US-0577409.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-514838/56.

XX N-PSDB; AAI83509.

XX Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders -

XX Claim 20; SEQ ID NO 17470; 1399pp + Sequence Listing; English.

XX The invention relates to human polynucleotides (AA179941-AA193841) and  
CC the encoded proteins (AA000010-AA013910) that exhibit activity relating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.

XX Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 104 AA;

Query Match 78.9%; Score 45; DB 22; Length 104;  
Best Local Similarity 85.7%; Pred. No. 24;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 PRWPPQ 8  
I:|||||  
Db 7 PQWPPQ 13

## RESULT 4

AAU30711  
ID AAU30711 standard; Protein; 195 AA.

XX AC AAU30711;

XX DT 18-DEC-2001 (first entry)

XX DE Novel human secreted protein #1202.

XX DE Human; vaccination; gene therapy; nutritional supplement;

KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;  
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.

XX OS Homo sapiens.

XX PN WO200179449-A2.

XX PD 25-OCT-2001.

XX PF 16-APR-2001; 2001WO-US08656.

XX PR 18-APR-2000; 2000US-0552929.

XX PR 26-JAN-2001; 2001US-0770160.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Drmanac RT;

XX DR WPI; 2001-611725/70.

XX PT Nucleic acids encoding a range of human polypeptides, useful in genetic  
PT vaccination, testing and therapy -

PS Claim 20; Page 334; 765pp; English.

XX CC The invention relates to novel human secreted polypeptides. The  
CC polypeptides and antibodies to the polypeptides are useful for  
CC determining the presence of or predisposition to a disease associated  
CC with altered levels of polypeptide. The polypeptides are also useful for  
CC identifying agents (agonists and antagonists) that bind to them. Cells  
CC expressing the proteins are useful for identifying a therapeutic agent  
CC for use in treatment of a pathology related to aberrant expression or  
CC physiological interactions of the polypeptide. Vectors comprising  
CC the nucleic acids encoding the polypeptides and cells genetically  
CC engineered to express them are also useful for producing the proteins.  
CC The proteins are useful in genetic vaccination, testing and  
CC therapy, and can be used as nutritional supplements. They may be used to  
CC increase stem cell proliferation; to regulate haematopoiesis; and in  
CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;  
CC immune suppression and/or stimulation; as anti-inflammatory agents; and  
CC in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid  
CC sequences of novel human secreted proteins of the invention.

SQ Sequence 195 AA;

Query Match 78.9%; Score 45; DB 22; Length 195;

Best Local Similarity 85.7%; Pred. No. 44;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 PRWPPQ 8

I:|||||

Db 49 PRWPPR 55

## RESULT 5

ABG08009

ID ABG08009 standard; Protein; 90 AA.

XX AC

ABG08009;

XX DT 13-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #8000.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-PSDB; AAS72196.

XX PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -

XX PS Claim 20; SEQ ID No 38368; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 90 AA;

Query Match 77.2%; Score 44; DB 22; Length 90;

Best Local Similarity 100.0%; Pred. No. 29;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 PRWPPP 7

I:|||||

Db 22 PRWPPP 27

## RESULT 6

AAU29704

ID AAU29704 standard; Protein; 443 AA.

XX AC

AAU29704;

```

XX DT 18-DEC-2001 (first entry)
XX DE Novel human secreted protein #195.
XX DE Human; vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX OS Homo sapiens.
XX PN WO200179449-A2.
XX PD 25-OCT-2001.
XX PF 16-APR-2001; 2001WO-US08656.
XX PR 18-APR-2000; 2000US-0552929.
XX PR 26-JAN-2001; 2001US-0770160.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Drmanac RT;
XX DR WPI; 2001-611725/70.
XX PT Nucleic acids encoding a range of human polypeptides, useful in genetic
XX vaccination, testing and therapy -
XX PS Claim 20; Page 182-183; 765pp; English.
XX CC The invention relates to novel human secreted polypeptides. The
XX polypeptides and antibodies to the polypeptides are useful for
XX determining the presence of or predisposition to a disease associated
XX with altered levels of polypeptide. The polypeptides are also useful for
XX identifying agents (agonists and antagonists) that bind to them. Cells
XX expressing the proteins are useful for identifying a therapeutic agent
XX for use in treatment of a pathology related to aberrant expression or
XX physiological interactions of the polypeptide. Vectors comprising
XX the nucleic acids encoding the polypeptides and cells genetically
XX engineered to express them are also useful for producing the proteins.
XX The proteins are useful in genetic vaccination, testing and
XX therapy, and can be used as nutritional supplements. They may be used to
XX increase stem cell proliferation; to regulate haematopoiesis; and in
XX bone, cartilage, tendon and/or nerve tissue growth or regeneration;
XX immune suppression and/or stimulation; as anti-inflammatory agents; and
XX in treatment of leukaemias. AAU29510-AAU3304 represent the amino acid
XX sequences of novel human secreted proteins of the invention.
XX SQ Sequence 443 AA;
Query Match 77.2%; Score 44; DB 22; Length 443;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 PRWPPP 7
DB 241 PRWPPP 246
II:IIII
RESULT 7
AAW58988
ID AAW58988 standard; Protein; 709 AA.
XX AC AAW58988;
XX DT 11-SEP-1998 (first entry)
XX DE Homo sapiens fetal kidney clone BD335_14 encoded protein.
XX KW fetal; kidney; cDNA library; clone BD335_14; secreted protein
KW autoimmune disease; anti-inflammatory; immune; stimulation
KW suppression.

```

```

XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT Misc-difference 255 /note= "undefined amino acid, encoded by AKC"
XX PN WO9814576-A2.
XX PD 09-APR-1998.
XX PF 03-OCT-1997; 97WO-US18007.
XX PR 04-OCT-1996; 96US-0726237.
XX PA (GEMY ) GENETICS INST INC.
XX PI Agostino MJ, Jacobs K, Lavallie ER, Mccoy JM, Marberg D;
XX PI Racie LA, Spaulding V, Treacy M;
XX DR WPI; 1998-240082/21.
XX DR N-PSDB; AAV11636.
XX PT Nucleic acids encoding novel secreted proteins - useful as, e.g.
XX anti-inflammatory, immuno-stimulatory or suppressing agents
XX PS Disclosure; Page 88-91; 110pp; English.
XX CC The sequence is that of a secreted protein encoded by
XX an isolated polynucleotide which may be of use in the
XX production of therapeutic compositions for treating or
XX ameliorating a medical condition in a mammal. Such compositions
XX may be used for, e.g. research purposes as markers for
XX tissues, molecular weight markers for gels, primers or probes, for
XX nutrition as carbon, nitrogen or carbohydrate source. They can also be
XX used as a cytokine for cell proliferation and differentiation activity,
XX as immune stimulants or suppressors, e.g. for viral, bacterial or fungal
XX infections, for autoimmune diseases such as multiple sclerosis or
XX systemic lupus erythematosus, to regulate haematopoiesis, for tissue
XX growth, as an activator or inhibitor, or as a chemotactic or
XX chemokinetic, haemostatic and thrombocytic receptor/ligand,
XX anti-inflammatory or tumour inhibitor agents.
XX SQ Sequence 709 AA;
Query Match 77.2%; Score 44; DB 19; Length 709;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 LPRWPPP 7
DB 135 LPQWPPP 141
II:IIII
RESULT 8
AAB90668
ID AAB90668 standard; Protein; 709 AA.
XX AC AAB90668;
XX DT 07-JUN-2001 (first entry)
XX DE Human BD335_14 protein sequence SEQ ID 5.
XX KW Human; secreted protein; nutrient; cytokine modulator; proliferation;
KW differentiation; immune system modulator; tissue growth; chemotactic;
KW haemostatic; thrombolytic; anti-inflammatory; tumour inhibition;
KW haematopoiesis.
XX OS Homo sapiens.
XX PN WO200119988-A1.
XX

```

PD 22-MAR-2001.  
 XX  
 PF 14-SEP-2000; 2000WO-US25135.  
 XX  
 PR 17-SEP-1999; 99US-0398829.  
 XX  
 PA (GEM) GENETICS INST INC.  
 XX  
 PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;  
 PI Merberg D, Treacy M, Bowman MR, Spaulding V, Agostino MJ;  
 XX  
 DR WPI; 2001-244801/25.  
 DR N-PSDB; AAF98376.  
 XX  
 PT Isolated nucleic acids encoding polypeptides, useful for modulating  
 PT e.g. cytokine and cell proliferation/differentiation activity, the  
 PT immune system and hematopoiesis regulating activity -  
 XX  
 PS Disclosure; Page 382-384; 557pp; English.  
 XX  
 CC Human cDNA clones represented in AAF98374 - AAF98489 encode secreted  
 CC proteins AAB90667 - AAB90750. The cDNA clones are isolated from various  
 CC tissue types, and may be used in the prevention, treatment and diagnosis  
 CC of diseases associated with inappropriate protein expression. The  
 CC polypeptides and nucleic acids may be used as nutrients or to modulate  
 CC cytokine and cell proliferation/differentiation activity and may also be  
 CC involved in modulation of the immune system. The cDNA sequences,  
 CC proteins, their agonists and/or antagonists exhibit hematopoiesis  
 CC regulating activity; tissue growth activity; activin/inhibin activity;  
 CC chemotactic/chemokinetic activity; haemostatic and thrombolytic  
 CC activity; receptor/ligand activity; anti-inflammatory activity;  
 CC haematopoiesis activity; cadherin/tumour suppressor activity; and/or  
 CC tumour inhibition activity. Included in the invention are probes  
 CC represented in AAF98490 - AAF98572 which are specific for the cDNA clones  
 CC encoding the secreted proteins.  
 XX  
 SQ Sequence 709 AA;  
 Query Match 77.2%; Score 44; DB 22; Length 709;  
 Best Local Similarity 85.7%; Pred. No. 2e+02; Mismatches 0; Indels 0; Gaps 0;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY . 1 LPRWPPP 7  
 Db 135 LPQWPPP 141  
 ||:||||  
 RESULT 9  
 ABB97356  
 ID ABB97356 standard; Protein; 1507 AA.  
 XX  
 AC ABB97356;  
 XX  
 DT 27-JUN-2002 (first entry)  
 XX  
 DE Novel human protein SEQ ID NO: 624.  
 XX  
 KW Human; antianaemic; vulnary; antiinflammatory; immunomodulator;  
 KW antiinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;  
 KW neuroprotective; antiparkinsonian; protein therapy; EST;  
 KW expressed sequence tag.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200222660-A2.  
 XX  
 PD 21-MAR-2002.  
 XX  
 PF 10-SEP-2001; 2001WO-US26015.  
 XX  
 PR 11-SEP-2000; 2000US-0659671.  
 XX  
 PA (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;  
 PI Xue AJ, Yang Y, Wehrman T, Dmanac RT;  
 XX  
 DR WPI; 2002-292408/33.  
 DR N-PSDB; ABN32542.  
 XX  
 PT An isolated polynucleotide for treating diseases associated with its  
 PT encoded polypeptide such as cancer and multiple sclerosis -  
 XX  
 PS Example 2; SEQ ID NO 624; 509pp; English.  
 XX  
 CC The present invention provides the protein and coding sequences of 444  
 CC novel human proteins. These were isolated from expressed sequences tags  
 CC (ESTs). They can be used to stimulate cell growth, to regulate  
 CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth  
 CC e.g. in burn treatment, to regulate the immune system e.g. to treat  
 CC multiple sclerosis, to regulate activin or inhibin e.g. to treat  
 CC infertility, to regulate haemostasis or thrombolysis e.g. to treat  
 CC stroke and cancer, to screen for drugs, to treat inflammatory conditions  
 CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.  
 CC Parkinson's disease. The present sequence is a protein of the invention.  
 XX  
 SQ Sequence 1507 AA;  
 Query Match 77.2%; Score 44; DB 23; Length 1507;  
 Best Local Similarity 85.7%; Pred. No. 3.9e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY . 1 LPRWPPP 7  
 Db 933 LPQWPPP 939  
 ||:||||  
 RESULT 10  
 AAR60620  
 ID AAR60620 standard; Protein; 1958 AA.  
 XX  
 AC AAR60620;  
 XX  
 DT 15-MAY-1995 (first entry)  
 XX  
 DE Protein from ORF2 of Pseudorabies virus large latency transcript.  
 XX  
 KW Pseudorabies virus; PRV; LLV; large latency transcript;  
 KW attenuated virus; vaccine; early protein 0; EP0; HSV-1 ICPO;  
 KW protecting animals; deletion mutants; swine; ds.  
 XX  
 OS Pseudorabies virus.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 330..333  
 FT /note="histidine rich region"  
 FT Region 363..368  
 FT /note="histidine rich region"  
 FT Region 381..384  
 FT /note="histidine rich region"  
 FT Region 547..573  
 FT /note="acidic residue rich region"  
 FT Region 1830..1833  
 FT /note="cysteine rich region"  
 FT Region 1837..1840  
 FT /note="cysteine rich region"  
 XX  
 PN US5352596-A.  
 XX  
 PD 04-OCT-1994.  
 XX  
 PF 11-SEP-1992; 92US-0945283.  
 XX  
 PR 11-SEP-1992; 92US-0945283.  
 XX  
 PA (USDA ) US SEC OF AGRIC.

XX Cheung AK, Wesley RD;  
 PI  
 XX WPI: 1994-316187/39.  
 DR N-PSDB; AAO73500.  
 XX  
 PT Novel pseudorabies virus mutants for use in vaccine - having a  
 PT deletion and/or insertion in the early protein O gene or large  
 PT latency transcript gene  
 XX  
 PS Disclosure: Column 15-30; 43pp; English.  
 XX  
 CC This protein sequence is encoded by ORF2 of LfT (large latency  
 CC transcript) from Pseudorabies virus (PRV) as shown in AAO73500.  
 CC The basic sequence is derived from PRV strain InFh and PRV  
 CC strain Ka. The LfT overlaps and is transcribed in the opposite  
 CC orientation with respect to the EP0 (early polypeptide 0) and the  
 CC immediately early gene (IE180). EP0 is nonessential for replicatio,  
 CC LfT is the only gene expressed during PRV latency, and the IE180  
 CC gene is absolutely necessary for PRV replication. However there are  
 CC 2 copies of IE180 in the genome. It is expected that PRV lacking one  
 CC of the IE180 copies is viable. Deletions in the non-overlapping  
 CC regions of these 3 genes will generate single deletion routants,  
 CC while deletions in overlapping regions will generate double deletion  
 CC mutants. The invention is concerned with the construction of attenuated  
 CC viruses which have a reduced ability to reactivate from latency. This  
 CC can be achieved by functionally disabling the expression of the EP0  
 CC gene, or by disrupting the synthesis of the LfT, or both. (See also  
 CC AAO73501 and AARG0620-24)  
 XX  
 SQ Sequence 1958 AA;  
 Query Match 77.2%; Score 44; DB 15; Length 1958;  
 Best Local Similarity 100.0%; Pred. No. 5e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 PRWPPP 7  
 | | | | |  
 DB 423 PRWPPP 428  
 RESULT 11  
 ID ABP09141  
 XX ABP09141 standard; Protein; 53 AA.  
 AC ABP09141;  
 XX  
 DT 24-JUN-2002 (first entry)  
 XX Human ORFX protein sequence SEQ ID NO:18264.  
 DE  
 XX Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;  
 KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;  
 KW degenerative disorder; osteoarthritis; neurodegenerative disorder;  
 KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
 KW hypertension; hypothyroidism; cholesterol ester storage disease;  
 KW immune deficiency; immune disorder; infectious disease;  
 KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
 KW myasthenia gravis.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200192523-A2.  
 XX  
 PD 06-DEC-2001.  
 XX  
 PF 29-MAY-2001; 2001WO-US10836.  
 XX  
 PR 30-MAY-2000; 2000US-206132P.  
 PR 29-AUG-2000; 2000US-228716P.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX

PI Shimkets RA, Leach MD;  
 XX  
 DR WPI: 2002-106308/14.  
 DR N-PSDB; ABN24893.  
 XX  
 PT Novel human polypeptides and polynucleotides useful for diagnosing,  
 PT preventing and treating cardiovascular disease, neurodegenerative,  
 PT hyperproliferative disorders and autoimmune disorders  
 XX  
 PS Disclosure: SEQ ID 18264; 1037pp; English.  
 XX  
 CC The present invention describes substantially purified human proteins  
 CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
 CC in the specification). ABN15762 to ABN27252 encode the human ORFX  
 CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for  
 CC treating or preventing a pathology associated with an ORFX-associated  
 CC disorder in humans, and in the manufacture of a medicament for treating a  
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
 CC sequences can be used in gene therapy. ORFX sequences can be used in the  
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
 CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,  
 CC osteoarthritis, neurodegenerative disorders, disorders related to organ  
 CC transplantation, cardiovascular diseases, diabetes mellitus, systemic  
 CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester  
 CC storage disease, various immune deficiencies and disorders, infectious  
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also  
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,  
 CC bone degenerative disorders, or periodontal disease, and for gut  
 CC protection or regeneration and treatment of lung or liver fibrosis,  
 CC reperfusion injury in various tissues and conditions resulting from  
 CC systemic cytokine damage.  
 CC N.B. The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 53 AA;  
 Query Match 75.4%; Score 43; DB 23; Length 53;  
 Best Local Similarity 85.7%; Pred. No. 25;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 PRWPPP 8  
 | | | | |  
 DB 46 PSWPPPQ 52  
 RESULT 12  
 ID AAO07833  
 XX AAO07833 standard; Protein; 75 AA.  
 AC AAO07833;  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX Human polypeptide SEQ ID NO 21725.  
 DE  
 XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorders; arthritis; inflammation.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200164835-A2.  
 XX  
 PD 07-SEP-2001.  
 XX  
 PF 26-FEB-2001; 2001WO-US04927.  
 XX  
 PR 28-FEB-2000; 2000US-0515126.  
 PR 18-MAY-2000; 2000US-0577409.



XX -PA (HYSE-) HYSEQ INC.  
XX PI Tang YT, Liu C, Drmanac RT;  
XX PS WPI: 2001-514838/56.  
XX DR N-PSDB: AAI87764.  
XX XX  
XX Isolated nucleic acids and polypeptides, useful for preventing  
XX diagnosing and treating e.g. leukaemia, inflammation and immune  
XX disorders -  
XX PS Claim 20; SEQ ID NO 21725; 1399pp + Sequence Listing; English.  
XX XX  
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and  
XX the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to  
XX cytokine, cell proliferation or cell differentiation or which may induce  
XX production of other cytokines in other cell populations. The  
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or  
XX peptide therapy. The polypeptides have various cytokine-like activities,  
XX e.g. stem cell growth factor activity, haematopoiesis regulating  
XX activity, tissue growth factor activity, immunomodulatory activity and  
XX activin/inhibin activity and may be useful in the diagnosis and/or  
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and  
XX inflammation.  
XX Note: The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.  
XX XX  
XX Sequence 75 AA;  
XX  
XX Query Match 75.4%; Score 43; DB 22; Length 75;  
XX Best Local Similarity 77.8%; Pred. No. 34;  
XX Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
XX  
XX QY 1 LPRWPPQL 9  
XX | | | | | | |  
XX Db 63 LPRWPPKAL 71  
XX  
XX RESULT 13  
XX ABG19493  
XX ID ABG19493 standard; Protein; 32 AA.  
XX AC ABG19493;  
XX DT 13-FEB-2002 (first entry)  
XX DE Novel human diagnostic protein #19484.  
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder.  
XX OS Homo sapiens.  
XX PN WO200175067-A2.  
XX PD 11-OCT-2001.  
XX PF 30-MAR-2001; 2001WO-US08631.  
XX PR 31-MAR-2000; 2000US-0540217.  
XX PR 23-AUG-2000; 2000US-0649167.  
XX XX  
XX (HYSE-) HYSEQ INC.  
XX PI Drmanac RT, Liu C, Tang YT;  
XX WPI: 2001-639362/73.  
XX DR N-PSDB: AAS93680.  
XX XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits and to assess  
XX biodiversity -

PT responsible for genetic disorders or other traits and to assess  
XX biodiversity -  
XX PS Claim 20; SEQ ID NO 49852; 103pp; English.  
XX XX  
XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving  
XX (II). (II) is useful for generating antibodies against it, detecting or  
XX quantitating a polypeptide in tissue, as molecular weight markers and as  
XX a food supplement. (II) and its binding partners are useful in medical  
XX disorders involving aberrant protein expression or biological activity.  
XX The polypeptide and polynucleotide sequences have applications in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits to assess biodiversity  
XX and to produce other types of data and products dependent on DNA and  
XX amino acid sequences. ABG00010-ABG30377 represent novel human  
XX diagnostic amino acid sequences of the invention.  
XX Note: The sequence data for this patent did not appear in the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.  
XX XX  
XX Sequence 32 AA;  
XX  
XX Query Match 74.6%; Score 42.5; DB 22; Length 32;  
XX Best Local Similarity 88.9%; Pred. No. 18;  
XX Matches 8; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
XX  
XX QY 2 PR-WPPQL 9  
XX | | | | | | |  
XX Db 13 PRWPPQL 21  
XX  
XX RESULT 14  
XX ABG02119  
XX ID ABG02119 standard; Protein; 142 AA.  
XX AC ABG02119;  
XX DT 13-FEB-2002 (first entry)  
XX DE Novel human diagnostic protein #2110.  
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder.  
XX OS Homo sapiens.  
XX PN WO200175067-A2.  
XX PD 11-OCT-2001.  
XX PF 30-MAR-2001; 2001WO-US08631.  
XX PR 31-MAR-2000; 2000US-0540217.  
XX PR 23-AUG-2000; 2000US-0649167.  
XX XX  
XX (HYSE-) HYSEQ INC.  
XX PI Drmanac RT, Liu C, Tang YT;  
XX WPI: 2001-639362/73.  
XX DR N-PSDB: AAS66306.  
XX XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits and to assess  
XX biodiversity -

```

XX PS Claim 20; SEQ ID NO 32478; 103pp; English.
XX CC
XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human
XX CC diagnostic amino acid sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 142 AA;

Query Match 73.7%; Score 42; DB 22; Length 142;
Best Local Similarity 85.7%; Pred. No. 85;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LPRWPPP 7
Db 126 LPTWPPP 132
|||
|

RESULT 15
ABG19033
ID ABG19033 standard; Protein; 429 AA.
AC ABG19033;
XX ABG19033;
XX 18-FEB-2002 (first entry)
XX Novel human diagnostic protein #19024.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI: 2001-639362/73.
XX N-PSDB; AAS83220.
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -
XX Claim 20; SEQ ID No 49392; 103pp; English.

```

```

XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human
XX CC diagnostic amino acid sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 429 AA;

Query Match 73.7%; Score 42; DB 22; Length 429;
Best Local Similarity 85.7%; Pred. No. 2.4e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LPRWPPP 7
Db 96 LPAWPPP 102
|||
|

Search completed: November 26, 2002, 12:01:26
Job time : 14.4839 secs

```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 26, 2002, 05:31:33 : Search time 9.87097 Seconds  
(without alignments)  
187.866 Million cell updates/sec

Title: US-09-674-593-3  
Perfect score: 57  
Sequence: 1 LPRWPPPPQL 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues  
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_21.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phase.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_rvirus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	57	100.0	84	Q9UBP8	Q9ubp8 homo sapien
2	44	77.2	156	Q9FTJ1	Q9ftj1 oryza sativ
3	44	77.2	767	Q96IR9	Q96ir9 homo sapien
4	44	77.2	1380	Q9P247	Q9p247 homo sapien
5	44	77.2	1958	Q96340	Q96340 pseudorabie
6	44	77.2	1995	Q8TF72	Q8tf72 homo sapien
7	43	75.4	642	Q9NSV4	Q9nsv4 caenorhabdi
8	42	73.7	211	Q9ZDV6	Q9zdv6 arabidopsis
9	42	73.7	769	Q52395	Q52395 mycobacteri
10	42	73.7	790	Q94GL4	Q94gl4 oryza sativ
11	42	73.7	2424	Q9VZ48	Q9vz48 drosophila
12	41	71.9	190	Q9RT47	Q9rt47 deinoecoccus
13	41	71.9	295	Q9SHP7	Q9shp7 arabidopsis
14	41	71.9	2785	Q8ZYB9	Q8zyb9 pyrobaculum
15	40	70.2	61	Q40658	Q40658 oryza sativ
16	40	70.2	166	Q8X310	Q8x310 escherichia

17	40	70.2	180	5	P90993	P90993 caenorhabdi
18	40	70.2	206	16	Q9AC33	Q9ac33 caulobacter
19	40	70.2	218	17	Q9YG73	Q9yg73 aeropyrum p
20	40	70.2	267	10	Q94DEL	Q94del oryza sativ
21	40	70.2	272	12	Q919W2	Q919m2 culex nigri
22	40	70.2	354	2	Q9REJ4	Q9rej4 bradyrhizob
23	40	70.2	409	5	Q9XUR1	Q9xur1 caenorhabdi
24	40	70.2	411	4	Q9BTQ4	Q9btq4 homo sapien
25	40	70.2	435	4	Q96DR0	Q96dr0 homo sapien
26	40	70.2	485	5	Q95RE0	Q95re0 drosophila
27	40	70.2	489	4	Q96E23	Q96e23 homo sapien
28	40	70.2	489	11	Q9BQC3	Q9bqc3 homo sapien
29	40	70.2	489	11	Q9CR25	Q9cr25 mus musculu
30	40	70.2	562	13	Q8UW75	Q8uw75 oryzias lat
31	40	70.2	571	5	O18357	O18357 drosophila
32	40	70.2	629	5	Q903E2	Q903e2 caenorhabdi
33	40	70.2	654	6	Q95LY1	Q95ly1 macaca fasc
34	40	70.2	654	6	Q95LV1	Q95lv1 macaca fasc
35	40	70.2	796	4	Q9H869	Q9h869 homo sapien
36	40	70.2	818	16	Q9F3E4	Q9f3e4 streptomyce
37	39	68.4	182	2	Q8RK85	Q8rkb5 fremyella d
38	39	68.4	221	10	Q9FN88	Q9fn88 arabidopsis
39	39	68.4	237	16	Q98ND6	Q98nd6 rhizobium l
40	39	68.4	274	4	Q96LW2	Q96lw2 homo sapien
41	39	68.4	299	16	Q8U8W0	Q8u8w0 agrobacteri
42	39	68.4	347	16	Q9K2U4	Q9kzu4 streptomyce
43	39	68.4	471	16	Q8X811	Q8xrl1 raietonia s
44	39	68.4	613	11	Q08841	Q08841 cavia porce
45	39	68.4	669	10	Q92WK0	Q92wk0 arabidopsis

## ALIGNMENTS

## RESULT 1

Q9UBP8 PRELIMINARY: PRT: 84 AA.  
 AC Q9UBP8;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE R2AS protein.  
 GN RU2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20069887; PubMed=10601354;  
 RA Van den Eynde B.J., Gaugler B., Probst-Keppler M., Michaux L.,  
 RA Devuyt O., Lorge F., Weynants P., Boon T.;  
 RT "A new antigen recognized by cytolytic T lymphocytes on a human kidney  
 tumor results from reverse strand transcription."  
 RL J. Exp. Med. 190:1793-1800(1999).  
 DR EMBL; AF181722; AAF23613.1;  
 DR EMBL; AF181720; AAF23611.1;  
 SQ SEQUENCE 84 AA, 8969 MW, 70B739F173A9E560 CRC64;

Query Match 100.0%; Score 57; DB 4; Length 84;  
 Best Local Similarity 100.0%; Pred. No. 0.037;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPRWPPPPQL 9  
 |||||  
 Db 38 LPRWPPPPQL 46

## RESULT 2

Q9FTJ1 PRELIMINARY: PRT: 156 AA.  
 ID Q9FTJ1;  
 AC Q9FTJ1;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE P0410E01.9 protein.  
GN P0410E01.9  
OS Oryza sativa (rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzaceae; Oryza.  
OX NCBI\_TaxID=4530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. NIPPONBARE;  
RA Sasaki T., Matsumoto T., Yamamoto K.;  
RT "Oryza sativa nipponbare(CA3) genomic DNA, chromosome 1, PAC  
clone: P0410E01.";  
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AP002866; BAB17088.1; -  
SQ SEQUENCE 156 AA; 16741 MW; 6677477714F4E3B6 CRC64;

Query Match 77.2%; Score 44; DB 10; Length 156;  
Best Local Similarity 77.8%; Pred. No. 7.7;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 LPRWPPPL 9  
DB 74 LPPWPPPL 82

## RESULT 3

Q96IR9 ID Q96IR9 PRELIMINARY; PRT; 767 AA.  
AC Q96IR9;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Similar to shroom (fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=COLON;  
RA Strausberg R.;  
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: BC007291; AAH07291.1; -  
FT NON\_TER 1  
SQ SEQUENCE 767 AA; 83291 MW; 1BF2520CD38C02B5 CRC64;

Query Match 77.2%; Score 44; DB 4; Length 767;  
Best Local Similarity 85.7%; Pred. No. 35;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPRWPPP 7  
DB 193 LPQWPPP 199

## RESULT 4

Q9P247 ID Q9P247 PRELIMINARY; PRT; 1380 AA.  
AC Q9P247;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE KIAA1481 protein (fragment).  
GN KIAA1481.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.

RX MEDLINE=20277482; PubMed=10819331;  
RA Nagase T., Kikuno R., Ishikawa K., Hirose M., Ohata O.;  
RT "Prediction of the coding sequences of unidentified human  
genes. XVII. The complete sequences of 100 new cDNA clones from brain  
which code for large proteins in vitro.";  
RL DNA Res. 7:143-150(2000).  
DR EMBL: AB040914; BAA96005.1; -  
FT NON\_TER 1  
SQ SEQUENCE 1380 AA; 149753 MW; 973465CFFB9BC458 CRC64;

Query Match 77.2%; Score 44; DB 4; Length 1380;  
Best Local Similarity 85.7%; Pred. No. 61;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPRWPPP 7  
DB 806 LPQWPPP 812

## RESULT 5

Q69340 ID Q69340 PRELIMINARY; PRT; 1958 AA.  
AC Q69340;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE ORF1, ORF2, and ORF3.  
OS Pseudorabies virus.  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Alphaherpesvirinae; Varicellovirus.  
OX NCBI\_TaxID=10345;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=INDIANA-FUNKHAUSER;  
RX MEDLINE=91374576; PubMed=1654441;  
RA Cheung A.K.;  
RT "Cloning of the latency gene and the early protein 0 gene of  
RT pseudorabies virus.";  
RL J. Virol. 65:5260-5271(1991).  
DR EMBL: M57505; AAA47468.1; -  
SQ SEQUENCE 1958 AA; 193474 MW; 0DEC36EDC29E2E68 CRC64;

Query Match 77.2%; Score 44; DB 12; Length 1958;  
Best Local Similarity 100.0%; Pred. No. 86;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 PRWPPP 7  
DB 423 PRWPPP 428

## RESULT 6

Q8TF72 ID Q8TF72 PRELIMINARY; PRT; 1995 AA.  
AC Q8TF72;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Shroom-related protein.  
GN HSHRML.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Koide M., Iio A., Obata K., Inagaki M., Yokota M., Ono T., Tuan R.S.;  
RT "Molecular cloning of FKBP12-associated protein.";  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AB055660; BAB84689.1; -  
SQ SEQUENCE 1995 AA; 216660 MW; D0914F4A034BAADA CRC64;

Query Match 77.2%; Score 44; DB 4; Length 1995;

```

Best Local Similarity 85.7%; Pred. No. 87;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPRWPPPP 7
DB 1421 LPQWPPPP 1427

RESULT 7
Q9NSV4
ID Q9NSV4 PRELIMINARY; PRT; 642 AA.
AC Q9NSV4;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 70.4 kDa protein.
GS C53D5.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Murray J., Rohlfing T., Morris M.;
RT "The sequence of C. elegans cosmid C53D5.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006624; AAF39787.1; -.
DR InterPro; IPR000101; Gglutrnspptase.
DR Pfam; PF01019; G-glu_transp; 1.
DR PRINTS; PR01210; GGTTRANSPTASE.
DR TIGRFAMS; TIGR00066; g_glu_trans; 1.
KW Hypothetical protein.
SQ SEQUENCE 642 AA; 70425 MW; 9D7601AB07110F3C CRC64;

Query Match 75.4%; Score 43; DB 5; Length 642;
Best Local Similarity 77.8%; Pred. No. 43;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 LPRWPPPPQ 9
DB 87 LPRWFGPSL 95

RESULT 8
Q9ZUV6
ID Q9ZUV6 PRELIMINARY; PRT; 211 AA.
AC Q9ZUV6;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE At2g28140 protein.
GS AT2G28140.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
thaliana.";
RL Nature 402:761-768(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA Lin X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC005851; AAC98452.1; -.
SQ SEQUENCE 211 AA; 24340 MW; B7DDF07BAC65BB89 CRC64;

Query Match 73.7%; Score 42; DB 10; Length 211;
Best Local Similarity 72.7%; Pred. No. 21;
Matches 8; Conservative 1; Mismatches 0; Indels 2; Gaps 1;

QY 1 LPRW--PPPQL 9
DB 156 LPQWRHPPQL 166

RESULT 9
O52395
ID O52395 PRELIMINARY; PRT; 769 AA.
AC O52395;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Cell division protein.
GN FTSH.
OS Mycobacterium smegmatis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1772;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SN2;
RA Anilkumar G., Ajitkumar P.;
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF037269; AAC32257.1; -.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003959; AAA_ATPase_cent.
DR InterPro; IPR003960; AAA_sub.
DR InterPro; IPR000642; Peptidase_M41.
DR Pfam; PF00004; AAA; 1.
DR Pfam; PF01434; Peptidase_M41; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRFAMS; TIGR01241; FlsH_fam; 1.
DR PROSITE; PS00674; AAA; 1.
KW ATP-binding.
SQ SEQUENCE 769 AA; 83649 MW; 738F37A17C78178E CRC64;

Query Match 73.7%; Score 42; DB 2; Length 769;
Best Local Similarity 85.7%; Pred. No. 73;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 PRWPPPPQ 8
DB 695 PGWPPPPQ 701

RESULT 10
Q94GL4
ID Q94GL4 PRELIMINARY; PRT; 790 AA.

```

AC Q94GL4;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Hypothetical protein.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoidae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NIPPONBARE;  
 RA de la Bastide M., Nascimento L., Spiegel L., Kirchoff K., Preston R.,  
 RA King L., Vil M.D., Baker J., Zutavern T., Santos L., Bell M.,  
 RA Miller B., Kuit K., Rodriguez S., Cunniss D.M., Balija V., Shah R.,  
 RA Bahret A., Bal H., Palmer L., Yang C., O'Shaughnessy A., Dedhia N.,  
 RA McCombie W.R.;  
 RT "Genomic Sequence for Oryza sativa, Nipponbare strain, clone  
 RT OSJNB0004114, from Chromosome 10."  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC090487; AAK92665.1;  
 SQ SEQUENCE 790 AA; 87798 MW; 245EB1FEEB28D55D CRC64;  
 Query Match 73.7%; Score 42; DB 10; Length 790;  
 Best Local Similarity 66.7%; Pred. No. 75;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 LPRWPPQL 9  
 Db '400 VPPWPPPSL 408  
 RESULT 11  
 Q9VZ48  
 ID Q9VZ48 PRELIMINARY; PRT; 2424 AA.  
 AC Q9VZ48;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE CG2174 protein.  
 GN CG2174.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Anantides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Dandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadenhead L.B., Davies P.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Flossler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulo G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003484; AAF47980.1;  
 DR HSSP: P08799; LMND.  
 DR FlyBase: FBgn0030252; CG2174.  
 DR InterPro: IPR000048; IQ\_region.  
 DR InterPro: IPR001609; myosin\_head.  
 DR InterPro: IPR000857; MYTH4.  
 DR Pfam: PF00612; IQ; 3.  
 DR Pfam: PF00063; myosin\_head; 1.  
 DR Pfam: PF00784; MYTH4; 1.  
 DR PRINTS: PR00193; MYOSINHEAVY.  
 DR ProDom: PD000355; myosin\_head; 1.  
 DR SMART: SM00015; IQ; 3.  
 DR SMART: SM00242; MYSC; 1.  
 DR SMART: SM00139; MYTH4; 1.  
 SQ SEQUENCE 2424 AA; 267616 MW; 8AD62AA33F9AA5D0 CRC64;  
 Query Match 73.7%; Score 42; DB 5; Length 2424;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 3 RWPAPPQ 8  
 Db 1714 RWPAPPQ 1719  
 RESULT 12  
 Q9RT47  
 ID Q9RT47 PRELIMINARY; PRT; 190 AA.  
 AC Q9RT47;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Hypothetical protein DR1919.  
 GN DR1919.  
 OS Deinococcus radiodurans.  
 OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;  
 OC Deinococcaceae; Deinococcus.  
 OX NCBI\_TaxID=1299;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=R1;  
 RX MEDLINE=20036896; PubMed=10567266;  
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
 RA Vamatheva J.J., Lam P., McDonald L., Utterback T., Zaleski C.,  
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
 RA Fraser C.M.;  
 RT "Genome sequence of the radioresistant bacterium Deinococcus  
 RT radiodurans R1."  
 RL Science 286:1571-1577(1999).  
 DR EMBL: AE002031; AAF11472.1;  
 DR TIGR: DR1919;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 190 AA; 21327 MW; 30326C4982CC4721 CRC64;

Query Match 71.9%; Score 41; DB 16; Length 190;  
Best Local Similarity 83.3%; Pred. No. 28;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 PRWPP 7  
|:||||  
Db 17 PKWPP 22

RESULT 13  
Q9SHP7 PRELIMINARY; PRT; 295 AA.  
AC Q9SHP7;  
DT 01-MAY-2000 (TReMBLrel. 13, Created)  
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE Flk23.15.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Chao Q., Shinn P., Brooks S., Buehler E., Dunn P., Khan S., Kim C.,  
RA Walker M., Brooks S., Altafi H., Araujo R., Conn L., Conway A.B.,  
RA Gonzalez A., Hansen N.F., Huizar L., Kremenetskaia I., Lenz C., Li J.,  
RA Liu S., Luos S., Rowley D., Schwartz J., Toriumi M., Vysotskaia V.,  
RA Yu G., Davis R.W., Federspiel N.A., Theologis A., Ecker J.R.;  
RT "Genomic sequence for Arabidopsis thaliana BAC Flk23 from chromosome  
I.";  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC007508; AAG22834.1; -;  
DR InterPro; IP8001087; Lipase\_GDSL.  
SQ SEQUENCE 295 AA; 32918 MW; 82F3ED7153BEB2EC CRC64;

Query Match 71.9%; Score 41; DB 10; Length 295;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LPRWPP 6  
|:||||  
Db .56 LPRWPP 61

RESULT 14  
Q8ZYB9 PRELIMINARY; PRT; 2785 AA.  
AC Q8ZYB9;  
DT 01-MAR-2002 (TReMBLrel. 20, Created)  
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)  
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)  
DE PARE2b.  
GN PAE0850.  
OS Pyrobaculum aerophilum.  
OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;  
OC Thermoproteaceae; Pyrobaculum.  
OX NCBI\_TaxID=13773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=IM2 / ATCC 51768 / DSM 7523;  
RX PubMed=11792869;  
RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,  
RA Miller J.H.;  
RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum  
aerophilum.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).  
DR EMBL; AE009790; AAL63076.1; -;  
KW Complete proteome.  
SQ SEQUENCE 2785 AA; 300989 MW; 1E144809E1D59C2D CRC64;

Query Match 71.9%; Score 41; DB 17; Length 2785;  
Best Local Similarity 85.7%; Pred. No. 3.6e+02;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 2 PRWPPQ 8  
|:||||  
Db 74 PRWPPQ 80

RESULT 15  
Q40658 PRELIMINARY; PRT; 61 AA.  
AC Q40658;  
DT 01-NOV-1996 (TReMBLrel. 01, Created)  
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
DT 01-JAN-1999 (TReMBLrel. 09, Last annotation update)  
DE Isocitrate dehydrogenase (Fragment).  
GN AK88.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzae; Oryza.  
OX NCBI\_TaxID=4530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94323471; PubMed=8049371;  
RA Umeda M., Hara C., Matsubayashi Y., Li H., Lui Q., Tadokoro F.,  
RA Aotsuka S., Uchimiya H.;  
RT "Expressed sequence tags from cultured cells of rice (Oryza sativa L.)  
under stressed conditions: analysis of transcripts of genes engaged in  
ATP-generating pathways.";  
RL Plant Mol. Biol. 25:469-478(1994).  
DR EMBL; D21069; BAA04645.1; -;  
FT NON\_TER 1 1  
FT NON\_TER 61 61  
SQ SEQUENCE 61 AA; 7204 MW; CA1655D36E56AIDE CRC64;

Query Match 70.2%; Score 40; DB 10; Length 61;  
Best Local Similarity 75.0%; Pred. No. 14;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 PRWPPQL 9  
|:||||  
Db 15 PAWPPWL 22

Search completed: November 26, 2002, 17:59:23  
Job time : 16.871 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 26, 2002, 12:00:06 ; Search time 3.19355 Seconds  
(without alignments)  
82.919 Million cell updates/sec

Title: US-09-674-593-3

Perfect score: 57

Sequence: 1 LPRWPPQL 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*

1: /cgn2\_6/ptodata/1/iaa/5A-COMB.pep.\*

2: /cgn2\_6/ptodata/1/iaa/5B-COMB.pep.\*

3: /cgn2\_6/ptodata/1/iaa/6A-COMB.pep.\*

4: /cgn2\_6/ptodata/1/iaa/6B-COMB.pep.\*

5: /cgn2\_6/ptodata/1/iaa/PCTUS-COMB.pep.\*

6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44	77.2	1958	1	US-07-945-283-2
2	39	68.4	78	1	US-08-487-359-5
3	39	68.4	78	1	US-08-222-798A-5
4	39	68.4	1481	2	US-08-616-844-40
5	39	68.4	1481	2	US-08-599-654-40
6	39	68.4	1481	3	US-08-944-868A-40
7	39	68.4	1481	3	US-08-944-423A-40
8	39	68.4	1481	3	US-08-944-496-40
9	38	66.7	492	4	US-09-724-864-39
10	37	64.9	10	1	US-08-255-272-1
11	37	64.9	357	5	PCT-US91-00899-14
12	37	64.9	369	2	US-08-424-224-2
13	37	64.9	369	2	PCT-US94-02891-69
14	37	64.9	405	1	US-07-914-281-8
15	37	64.9	405	1	US-08-393-246-8
16	37	64.9	405	1	US-08-525-058A-8
17	37	64.9	405	2	US-08-483-151-4
18	37	64.9	405	2	US-08-696-731-8
19	37	64.9	405	4	US-09-042-531-8
20	37	64.9	577	4	US-09-413-814-4
21	37	64.9	1001	4	US-08-884-569A-2
22	37	64.9	1130	2	US-08-519-547A-6
23	36	63.2	20	4	US-08-602-999A-128
24	36	63.2	20	4	US-08-630-915A-181
25	36	63.2	20	4	US-08-630-915A-223
26	36	63.2	20	4	US-09-500-124-128
27	36	63.2	24	4	US-08-630-915A-41

28	36	63.2	79	1	US-08-487-359-2	Sequence 2, Appli
29	36	63.2	79	1	US-08-487-359-6	Sequence 6, Appli
30	36	63.2	79	1	US-08-222-798A-2	Sequence 2, Appli
31	36	63.2	79	1	US-08-222-798A-6	Sequence 6, Appli
32	36	63.2	170	4	US-08-259-451-19	Sequence 19, Appli
33	36	63.2	356	4	US-09-092-315-12	Sequence 12, Appli
34	36	63.2	393	4	US-09-390-131-8	Sequence 8, Appli
35	36	63.2	2289	3	US-09-051-019-2	Sequence 2, Appli
36	35.5	62.3	129	4	US-09-199-637A-97	Sequence 97, Appli
37	35	61.4	13	4	US-08-602-999A-80	Sequence 80, Appli
38	35	61.4	13	4	US-08-278-865-80	Sequence 80, Appli
39	35	61.4	13	4	US-09-500-124-80	Sequence 80, Appli
40	35	61.4	15	4	US-08-602-999A-342	Sequence 342, App
41	35	61.4	15	4	US-09-500-124-342	Sequence 342, App
42	35	61.4	45	4	US-08-602-999A-48	Sequence 48, Appli
43	35	61.4	45	4	US-08-278-865-48	Sequence 48, Appli
44	35	61.4	45	4	US-09-500-124-48	Sequence 48, Appli
45	35	61.4	80	4	US-09-134-001C-2843	Sequence 2843, Ap

#### ALIGNMENTS

#### RESULT 1

US-07-945-283-2

; Sequence 2, Application US/07945283

; Patent No. 5352596

; GENERAL INFORMATION:

; APPLICANT: Cheung, Andrew K.

; APPLICANT: Wesley, Ronald D.

; TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants

; TITLE OF INVENTION: Involving The EP0 and LLT Genes

; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Curtis P. Ribando

; STREET: 1815 No. 5352596th University Street

; CITY: Peoria

; STATE: IL

; COUNTRY: USA

; ZIP: 61604

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA: US/07/945,283

; APPLICATION NUMBER: US/07/945,283

; FILING DATE: 19920911

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: Ribando, Curtis P

; REGISTRATION NUMBER: 27976

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 309-685-4011 ext.513

; TELEFAX: 309-685-4128

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1958 amino acids

; TYPE: AMINO ACID

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-07-945-283-2

Query Match 77.2%; Score 44; DB 1; Length 1958;

Best Local Similarity 100.0%; Pred. No. 1.le+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 PRWPPP 7

Db 423 PRWPPP 428

#### RESULT 2

US-08-487-359-5  
; Sequence 5, Application US/08487359  
; Patent No. 5633229  
; GENERAL INFORMATION:  
; APPLICANT: KOKRYAKOV, VLADIMIR N.  
; APPLICANT: HARMIG, SYLVIA S.L.  
; APPLICANT: LEHRER, ROBERT I.  
; TITLE OF INVENTION: PROPHENINS-ANTIBIOTIC PEPTIDES  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 Pennsylvania Ave. N.W.  
; CITY: Washington, D.C.  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 20006-1812  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/487,359  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/222,798  
; FILING DATE: 05-APR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 2000-0553.00  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 887-0763  
; TELEX: 90-4030  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 78 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-487-359-5

Query Match 68.4%; Score 39; DB 1; Length 78;  
Best Local Similarity 85.7%; Pred. No. 25;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 PRWPPQ 8  
Db 30 PRFPPQ 36

RESULT 3  
US-08-222-798A-5  
; Sequence 5, Application US/08222798A  
; Patent No. 5804553  
; GENERAL INFORMATION:  
; APPLICANT: KOKRYAKOV, VLADIMIR N.  
; APPLICANT: HARMIG, SYLVIA S.L.  
; APPLICANT: LEHRER, ROBERT I.  
; TITLE OF INVENTION: PROPHENINS-ANTIBIOTIC PEPTIDES  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 Pennsylvania Ave. N.W.  
; CITY: Washington, D.C.  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 20006-1812  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/222,798A  
; FILING DATE: 05-APR-1994  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 2000-0553.00  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 887-0763  
; TELEX: 90-4030  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 78 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-222-798A-5

Query Match 68.4%; Score 39; DB 1; Length 78;  
Best Local Similarity 85.7%; Pred. No. 25;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 PRWPPQ 8  
Db 30 PRFPPQ 36

RESULT 4  
US-08-616-844-40  
; Sequence 40, Application US/08616844  
; Patent No. 5849578  
; GENERAL INFORMATION:  
; APPLICANT: FALB, DEAN A.  
; TITLE OF INVENTION: COMPOSITION AND METHODS FOR THE  
; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE  
; NUMBER OF SEQUENCES: 54  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/616,844  
; FILING DATE: 15-MAR-1996  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/599,654  
; FILING DATE: 09-FEB-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/485,573  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/386,844  
; FILING DATE: 10-FEB-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CORUZZI, LAURA A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7853-053  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-8864  
; TELEX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 40:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1481 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-616-844-40

Query Match 68.4%; Score 39; DB 2; Length 1481;  
Best Local Similarity 85.7%; Pred. No. 4.3e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 RWPPPL 9  
| | | | |  
Db 8 RWPPPL 14

## RESULT 5

US-08-599-654-40  
; Sequence 40, Application US/08599654  
; Patent No. 5882925  
; GENERAL INFORMATION:

; APPLICANT: FALB, DEAN A  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE  
; NUMBER OF SEQUENCES: 54  
; CORRESPONDENCE ADDRESS:

ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/599,654

FILING DATE: 09-FEB-1996

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/485,573

FILING DATE: 07-JUN-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/386,844

FILING DATE: 10-FEB-1995

ATTORNEY/AGENT INFORMATION:

NAME: CORUZZI, LAURA A

REGISTRATION NUMBER: 30,742

REFERENCE/DOCKET NUMBER: 7853-041

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-8864

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 40:

SEQUENCE CHARACTERISTICS:

LENGTH: 1481 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: unknown

MOLECULE TYPE: protein

US-08-599-654-40

Query Match 68.4%; Score 39; DB 2; Length 1481;  
Best Local Similarity 85.7%; Pred. No. 4.3e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 RWPPPL 9  
| | | | |  
Db 8 RWPPPL 14

## RESULT 6

US-08-944-868A-40  
; Sequence 40, Application US/08944868A  
; Patent No. 6018025  
; GENERAL INFORMATION:

APPLICANT: FALB, DEAN A

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE

TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE

NUMBER OF SEQUENCES: 54

CORRESPONDENCE ADDRESS:

ADDRESSEE: PENNIE & EDMONDS

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/944,868A

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/599,654

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/386,844

FILING DATE: 10-FEB-1995

ATTORNEY/AGENT INFORMATION:

NAME: CORUZZI, LAURA A

REGISTRATION NUMBER: 30,742

REFERENCE/DOCKET NUMBER: 7853-041

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-8864

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 40:

SEQUENCE CHARACTERISTICS:

LENGTH: 1481 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: unknown

MOLECULE TYPE: protein

US-08-944-868A-40

Query Match 68.4%; Score 39; DB 3; Length 1481;

Best Local Similarity 85.7%; Pred. No. 4.3e+02;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 RWPPPL 9

| | | | |

Db 8 RWPPPL 14

## RESULT 7

US-08-944-423A-40  
; Sequence 40, Application US/08944423A  
; Patent No. 6020463  
; GENERAL INFORMATION:

APPLICANT: FALB, DEAN A

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE

TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE

NUMBER OF SEQUENCES: 54

CORRESPONDENCE ADDRESS:

ADDRESSEE: PENNIE & EDMONDS

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

```
;
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: WINDOWS 95
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944,423A
; FILING DATE: 06-OCT-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/599,654
; FILING DATE: 09-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,573
; FILING DATE: JUN-07-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/386,844
; FILING DATE: 10-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: CORUZZI, LAURA A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-105
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1481 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-944-423A-40

Query Match 68.4%; Score 39; DB 3; Length 1481;
Best Local Similarity 85.7%; Pred. No. 4.3e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 RWPPPOL 9
Db 8 RWPPPLL 14

RESULT 8
US-08-944-496-40
; Sequence 40, Application US/08944496
; Patent No. 612433
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITL OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944,496
; FILING DATE: 06-OCT-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/599,654
```

```
;
; FILING DATE: 09-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,573
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/386,844
; FILING DATE: 10-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: CORUZZI, LAURA A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-104
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1481 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-944-496-40

Query Match 68.4%; Score 39; DB 3; Length 1481;
Best Local Similarity 85.7%; Pred. No. 4.3e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 RWPPPOL 9
Db 8 RWPPPLL 14

RESULT 9
US-09-724-864-39
; Sequence 39, Application US/09724864
; Patent No. 6380362
; GENERAL INFORMATION:
; APPLICANT: Watson, James D
; APPLICANT: Murlison, James G
; TITLE OF INVENTION: Polynucleotides, polypeptides expressed
; FILE REFERENCE: 11000.105001
; CURRENT APPLICATION NUMBER: US/09/724,864
; CURRENT FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: U.S. No. 6380362 60/171,678
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 39
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Mouse
; US-09-724-864-39

Query Match 66.7%; Score 38; DB 4; Length 492;
Best Local Similarity 83.3%; Pred. No. 2.1e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 PRWPPP 7
Db 425 PTWPPP 430

RESULT 10
US-08-253-272-1
; Sequence 1, Application US/08252572
; Patent No. 5627268
; GENERAL INFORMATION:
; APPLICANT: Kumar, Ramesh
; APPLICANT: Sharma, Ajay
; APPLICANT: Khoury-Christianson, Anastasia
; APPLICANT: M.
```

;; TITLE OF INVENTION: Production of Therapeutic Peptides in  
;; TITLE OF INVENTION: Transgenic Animals as a Fusion with Hemoglobin  
;; NUMBER OF SEQUENCES: 32  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: PENNIE & EDMONDS  
;; STREET: 1155 Avenue of the Americas  
;; CITY: New York  
;; STATE: NY  
;; COUNTRY: USA  
;; ZIP: 10036-2711  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/255,272  
;; FILING DATE:  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Coruzzi, Laura A.  
;; REGISTRATION NUMBER: 30742  
;; REFERENCE/DOCKET NUMBER: 6794-032  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (212) 790-9090  
;; TELEFAX: (212) 869-9741/8864  
;; INFORMATION FOR SEQ ID NO: 1:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 10 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: unknown  
;; MOLECULE TYPE: peptide  
;; HYPOTHETICAL: NO  
;; ANTI-SENSE: NO  
;; US-08-255-272-1

Query Match 64.9%; Score 37; DB 1; Length 10;  
Best Local Similarity 62.5%; Pred. No. 6.4;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 PRWPPQL 9  
Db 1 PWRPQI 8

RESULT 11  
PCT-US91-00899-14  
; Sequence 14, Application PC/TUS9100899  
; GENERAL INFORMATION:  
; APPLICANT: Lowe, John B.  
; TITLE OF INVENTION: Method and Products For the Synthesis of  
; TITLE OF INVENTION: Oligosaccharide Structures on Glycoproteins, Glycolipids,  
; TITLE OF INVENTION: or as Free Molecules, and for the Isolation of Cloned  
; TITLE OF INVENTION: Genetic Sequences That Determine These Structures  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
; STREET: 1755 Jefferson Davis Highway, Suite 400  
; CITY: Arlington  
; STATE: Virginia  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US91/00899  
; FILING DATE: 19910214  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lavalleye Ph.D., Jean-Paul

;; REGISTRATION NUMBER: 31,451  
;; REFERENCE/DOCKET NUMBER: 2363-021-55 PCT  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (703)521-5940  
;; TELEFAX: (703)486-2347  
;; TELEX: 248855 OPAT UR  
;; INFORMATION FOR SEQ ID NO: 14:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 357 amino acids  
;; TYPE: AMINO ACID  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; FRAGMENT TYPE: C-terminal  
;; PCT-US91-00899-14

Query Match 64.9%; Score 37; DB 5; Length 357;  
Best Local Similarity 83.3%; Pred. No. 2.1e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 PRWPPP 7  
Db 71 PDWPPP 76

RESULT 12  
US-08-424-224-2  
; Sequence 2, Application US/08424224  
; Patent No. 5912173  
; GENERAL INFORMATION:  
; APPLICANT: LEONARD, WARREN J.  
; TITLE OF INVENTION: MURINE IL-2R CDNA AND  
; TITLE OF INVENTION: USES THEREOF  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORGAN & FINNEGAN  
; STREET: 345 PARK AVE.  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WORD PERFECT # 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/424,224  
; FILING DATE:  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/121,435  
; FILING DATE: 14-SEPT-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: WILLIAM S. FEILER  
; REGISTRATION NUMBER: 26,728  
; REFERENCE/DOCKET NUMBER: 2026-4061US1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-758-4800  
; TELEFAX: 212-751-6849  
; TELEX: 421792  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 369  
; TYPE: AMINO ACID  
; TOPOLOGY: UNKNOWN  
; MOLECULE TYPE:  
; DESCRIPTION: PROTEIN  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: MURINE  
; INDIVIDUAL ISOLATE: IL-2R  
; US-08-424-224-2

Query Match 64.9%; Score 37; DB 2; Length 369;  
Best Local Similarity 83.3%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 PRWPPP 7  
| | | | |  
DB 356 PYWPPP 361

## RESULT 13

PCT-US94-02891-69

; Sequence 69, Application PC/TUS9402891  
; GENERAL INFORMATION:  
; APPLICANT: THE GOVERNMENT OF THE UNITED STATES OF AMERICA AS  
; APPLICANT: REPRESENTED BY THE SECRETARY, DEPARTMENT OF HEALTH AND HUMAN  
; APPLICANT: SERVICES  
; APPLICANT: OFFICE OF TECHNOLOGY TRANSFER, NATIONAL  
; APPLICANT: INSTITUTES OF HEALTH, BOX 017, BETHESDA, MARYLAND 20892 USA  
; TITLE OF INVENTION: METHODS FOR DIAGNOSIS AND TREATMENT OF  
; TITLE OF INVENTION: XSCID  
; NUMBER OF SEQUENCES: 69  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORGAN & FINNEGAN  
; STREET: 345 PARK AVE.  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: USA  
; ZIP: 10154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WORD PERFECT # 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/02891  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/031,143  
; FILING DATE: 12-MAR-1993  
; APPLICATION NUMBER: 08/121,435  
; FILING DATE: 14-SEPT-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: WILLIAM S. FEILER  
; REGISTRATION NUMBER: 26,728  
; REFERENCE/DOCKET NUMBER: 2026-4061  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-758-4800  
; TELEFAX: 212-751-6849  
; TELEX: 421792  
; INFORMATION FOR SEQ. ID NO: 69:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 369  
; TYPE: AMINO ACID  
; TOPOLOGY: UNKNOWN  
; MOLECULE TYPE:  
; DESCRIPTION: PROTEIN  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: MURINE  
; INDIVIDUAL ISOLATE: IL-2R  
; PCT-US94-02891-69

Query Match 64.9%; Score 37; DB 5; Length 369;  
Best Local Similarity 83.3%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 PRWPPP 7  
| | | | |  
DB 356 PYWPPP 361

## RESULT 14

US-07-914-281-8

; Sequence 8, Application US/07914281  
; Patent No. 5324663

; GENERAL INFORMATION:

; APPLICANT: LOWE, JOHN B.

; TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS

; TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS, ISOLATION

; TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION

; TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU

; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER &amp; NEUSTADT,

; ADDRESSEE: P.C.

; STREET: 1755 Jefferson Davis Highway, Fourth Floor

; CITY: Arlington

; STATE: Virginia

; COUNTRY: U.S.A.

; ZIP: 22202

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/914,281

; FILING DATE: 19920720

; CLASSIFICATION: 530

; ATTORNEY/AGENT INFORMATION:

; NAME: Lavalleye, Jean-Paul M. P.

; REGISTRATION NUMBER: 31,451

; REFERENCE/DOCKET NUMBER: 2363-060-55

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703)521-4500

; TELEFAX: (703)486-2347

; TELEX: 248855 OPAT UR

; INFORMATION FOR SEQ. ID NO: 8:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 405 amino acids

; TYPE: AMINO ACID

; TOPOLOGY: unknown

; MOLECULE TYPE: protein

; US-07-914-281-8

Query Match 64.9%; Score 37; DB 1; Length 405;  
Best Local Similarity 83.3%; Pred. No. 2.4e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 PRWPPP 7  
| | | | |  
DB 122 PDWPPP 127

## RESULT 15

US-08-393-246-8

; Sequence 8, Application US/08393246

; Patent No. 5595900

; GENERAL INFORMATION:

; APPLICANT: LOWE, JOHN B.

; TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS

; TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,

; TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION

; TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU

; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER &amp; NEUSTADT,

; ADDRESSEE: P.C.

; STREET: 1755 Jefferson Davis Highway, Fourth Floor

; CITY: Arlington

; STATE: Virginia

; COUNTRY: U.S.A.

; ZIP: 22202

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/393,246  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/220,433  
FILING DATE: 30-MAR-1994  
APPLICATION NUMBER: US 07/914,281  
FILING DATE: 20-JUL-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Lavalleye, Jean-Paul M. P.  
REGISTRATION NUMBER: 31,451  
REFERENCE/DOCKET NUMBER: 2363-060-55  
TELEPHONE: (703)521-4500  
TELEFAX: (703)486-2347  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 405 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-393-246-8

Query Match 64.9%; Score 37; DB 1; Length 405;  
Best Local Similarity 83.3%; Pred. No. 2.4e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 PRWPPP 7  
| | | | |  
Db 122 PDWPPP 127

Search completed: November 26, 2002, 19:43:20  
Job time : 5.19355 secs

**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 26, 2002, 12:01:36 ; Search time 4.74194 Seconds  
(without alignments)  
30.223 Million cell updates/sec

Title: US-09-674-593-3

Perfect score: 57

Sequence: 1 LPRWPPQL 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 102317 seqs, 15924203 residues

Total number of hits satisfying chosen parameters: 102317

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published\_Applications\_AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40	70.2	654	9	US-09-905-291A-177
2	40	70.2	654	10	US-09-909-320-177
3	40	70.2	654	10	US-09-909-0888-177
4	40	70.2	654	12	US-10-052-586-26
5	39	68.4	43	10	US-09-864-761-44755
6	39	68.4	229	9	US-09-764-868-831
7	39	68.4	1481	10	US-09-371-900-40
8	39	68.4	1481	10	US-09-924-417-60
9	38	66.7	71	10	US-09-864-761-42815
10	38	66.7	348	10	US-09-764-898-188
11	38	66.7	482	10	US-09-764-853-568
12	38	66.7	1198	10	US-09-866-582-36
13	38	66.7	1241	12	US-10-001-215-5
14	37	64.9	39	10	US-09-864-761-46140
15	37	64.9	52	10	US-09-864-761-40064
16	37	64.9	79	10	US-09-925-300-1002
17	37	64.9	83	10	US-09-790-264-17
18	37	64.9	103	10	US-09-764-864-1448
19	37	64.9	160	10	US-09-764-864-1020

Sequence 553, App  
Sequence 23, Appl  
Sequence 20, Appl  
Sequence 4, Appl  
Sequence 4, Appl  
Sequence 4, Appl  
Sequence 218, App  
Sequence 841, App  
Sequence 290, App  
Sequence 670, App  
Sequence 2, Appl  
Sequence 2, Appl  
Sequence 2, Appl  
Sequence 1817, Ap  
Sequence 1817, Ap  
Sequence 436, App  
Sequence 1127, Ap  
Sequence 12, Appl  
Sequence 90, App  
Sequence 8, Appl  
Sequence 19, Appl  
Sequence 17, Appl  
Sequence 6, Appl  
Sequence 18, Appl  
Sequence 16, Appl

102 64.9 184 10 US-09-925-297-553  
102 64.9 210 10 US-09-790-264-23  
22 37 64.9 247 10 US-09-790-264-20  
23 37 64.9 252 10 US-09-738-396-4  
24 37 64.9 252 10 US-09-738-396-4  
25 37 64.9 252 10 US-09-771-961-4  
26 37 64.9 287 10 US-09-764-898-218  
27 37 64.9 288 10 US-09-764-853-841  
28 37 64.9 288 10 US-09-764-898-290  
29 37 64.9 314 10 US-09-764-853-670  
30 37 64.9 327 10 US-09-738-396-2  
31 37 64.9 327 10 US-09-773-774A-2  
32 37 64.9 327 10 US-09-771-961-2  
33 37 64.9 357 9 US-09-736-457-1817  
34 37 64.9 357 9 US-09-902-941-1817  
35 37 64.9 357 10 US-09-833-790-436  
36 37 64.9 359 10 US-09-925-301-1127  
37 64.9 369 9 US-09-895-593-12  
38 37 64.9 369 10 US-09-895-943-12  
39 37 64.9 393 9 US-09-764-868-690  
40 37 64.9 405 10 US-09-863-475A-8  
41 37 64.9 830 10 US-09-064-199-19  
42 37 64.9 1106 10 US-09-064-199-17  
43 37 64.9 1130 9 US-10-104-595-6  
44 37 64.9 1130 10 US-09-064-199-18  
45 37 64.9 1207 10 US-09-064-199-16

#### ALIGNMENTS

#### RESULT 1

US-09-905-291A-177  
; Sequence 177, Application US/09905291A  
; Patent No. US20020160374A1  
; GENERAL INFORMATION:  
; APPLICANT: Genentech, Inc.  
; APPLICANT: Ashkenazi, Avi  
; APPLICANT: Botstein, David  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Ferrara, Napoleone  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Fong, Sherman  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerber, Hanspeter  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, A.  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, Christopher J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Hillan, Kenneth, J.  
; APPLICANT: Kljavin, Ivar J.  
; APPLICANT: Mather, Jennie P.  
; APPLICANT: Pan, James  
; APPLICANT: Paoni, Nicholas F.  
; APPLICANT: Roy, Margaret Ann  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Williams, P. Mickey  
; APPLICANT: Wood, William, I.  
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
; FILE REFERENCE: 10466-14  
; CURRENT APPLICATION NUMBER: US/09/905,291A  
; PRIOR FILING DATE: 2001-07-12  
; PRIOR APPLICATION NUMBER: PCT/US00/04414  
; PRIOR FILING DATE: 2000-02-22  
; PRIOR APPLICATION NUMBER: US 60/143,048  
; PRIOR FILING DATE: 1999-07-07  
; PRIOR APPLICATION NUMBER: US 60/145,698  
; PRIOR FILING DATE: 1999-07-26  
; PRIOR APPLICATION NUMBER: US 60/146,222

;; PRIOR FILING DATE: 1999-07-28  
;; PRIOR APPLICATION NUMBER: PCT/US99/20594  
;; PRIOR FILING DATE: 1999-09-08  
;; PRIOR APPLICATION NUMBER: PCT/US99/20944  
;; PRIOR FILING DATE: 1999-09-13  
;; PRIOR APPLICATION NUMBER: PCT/US99/21090  
;; PRIOR FILING DATE: 1999-09-15  
;; PRIOR APPLICATION NUMBER: PCT/US99/21547  
;; PRIOR FILING DATE: 1999-09-15  
;; PRIOR APPLICATION NUMBER: PCT/US99/23089  
;; PRIOR FILING DATE: 1999-10-05  
;; PRIOR APPLICATION NUMBER: PCT/US99/28214  
;; PRIOR FILING DATE: 1999-11-29  
;; PRIOR APPLICATION NUMBER: PCT/US99/28313  
;; PRIOR FILING DATE: 1999-11-30  
;; PRIOR APPLICATION NUMBER: PCT/US99/28564  
;; PRIOR FILING DATE: 1999-12-02  
;; PRIOR APPLICATION NUMBER: PCT/US99/28565  
;; PRIOR FILING DATE: 1999-12-02  
;; PRIOR APPLICATION NUMBER: PCT/US99/30095  
;; PRIOR FILING DATE: 1999-12-16  
;; PRIOR APPLICATION NUMBER: PCT/US99/30911  
;; PRIOR FILING DATE: 1999-12-20  
;; PRIOR APPLICATION NUMBER: PCT/US99/30999  
;; PRIOR FILING DATE: 1999-12-20  
;; PRIOR APPLICATION NUMBER: PCT/US00/00219  
;; PRIOR FILING DATE: 2000-01-05  
;; NUMBER OF SEQ ID NOS: 423  
;; SEQ ID NO 177  
;; LENGTH: 654  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-09-905-291A-177

Query Match 70.2%; Score 40; DB 9; Length 654;  
Best Local Similarity 75.0%; Pred. No. 1.1e+02;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LPRWPPQ 8  
11:11111  
Db 526 LPKWPYPQ 533

## RESULT 2

US-09-909-320-177  
; Sequence 177, Application US/09909320  
; Patent No. US20020132240A1  
; GENERAL INFORMATION:  
; APPLICANT: Genentech, Inc.  
; APPLICANT: Ashkenazi, Avi  
; APPLICANT: Botstein, David  
; APPLICANT: Desnovers, Luc  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Ferrara, Napoleone  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Fong, Sherman  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerber, Hanspeter  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, A.  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, Christopher J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Hillan, Kenneth, J.  
; APPLICANT: Kijavlin, Ivar J.  
; APPLICANT: Mather, Jennie P.  
; APPLICANT: Pan, James  
; APPLICANT: Paoni, Nicholas F.  
; APPLICANT: Roy, Margaret Ann  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Williams, P. Mickey  
; APPLICANT: Wood, William, I.

;; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
;; FILE REFERENCE: 10466-14  
;; CURRENT APPLICATION NUMBER: US/09/909,320  
;; PRIOR FILING DATE: 2002-01-04  
;; PRIOR APPLICATION NUMBER: PCT/US00/04414  
;; PRIOR FILING DATE: 2000-02-22  
;; PRIOR APPLICATION NUMBER: US 60/143,048  
;; PRIOR FILING DATE: 1999-07-07  
;; PRIOR APPLICATION NUMBER: US 60/145,698  
;; PRIOR FILING DATE: 1999-07-26  
;; PRIOR APPLICATION NUMBER: US 60/146,222  
;; PRIOR FILING DATE: 1999-07-28  
;; PRIOR APPLICATION NUMBER: PCT/US99/20594  
;; PRIOR FILING DATE: 1999-09-08  
;; PRIOR APPLICATION NUMBER: PCT/US99/20944  
;; PRIOR FILING DATE: 1999-09-13  
;; PRIOR APPLICATION NUMBER: PCT/US99/21090  
;; PRIOR FILING DATE: 1999-09-15  
;; PRIOR APPLICATION NUMBER: PCT/US99/21547  
;; PRIOR FILING DATE: 1999-09-15  
;; PRIOR APPLICATION NUMBER: PCT/US99/23089  
;; PRIOR FILING DATE: 1999-10-05  
;; PRIOR APPLICATION NUMBER: PCT/US99/28214  
;; PRIOR FILING DATE: 1999-11-29  
;; PRIOR APPLICATION NUMBER: PCT/US99/28313  
;; PRIOR FILING DATE: 1999-11-30  
;; PRIOR APPLICATION NUMBER: PCT/US99/28564  
;; PRIOR FILING DATE: 1999-12-02  
;; PRIOR APPLICATION NUMBER: PCT/US99/28565  
;; PRIOR FILING DATE: 1999-12-02  
;; PRIOR APPLICATION NUMBER: PCT/US99/30095  
;; PRIOR FILING DATE: 1999-12-16  
;; PRIOR APPLICATION NUMBER: PCT/US99/30911  
;; PRIOR FILING DATE: 1999-12-20  
;; PRIOR APPLICATION NUMBER: PCT/US99/30999  
;; PRIOR FILING DATE: 1999-12-20  
;; PRIOR APPLICATION NUMBER: PCT/US00/00219  
;; PRIOR FILING DATE: 2000-01-05  
;; NUMBER OF SEQ ID NOS: 423  
;; SEQ ID NO 177  
;; LENGTH: 654  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-09-909-320-177

Query Match 70.2%; Score 40; DB 10; Length 654;  
Best Local Similarity 75.0%; Pred. No. 1.1e+02;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LPRWPPQ 8  
11:11111  
Db 526 LPKWPYPQ 533

## RESULT 3

US-09-909-088B-177  
; Sequence 177, Application US/09909088B  
; Patent No. US20020146709A1  
; GENERAL INFORMATION:  
; APPLICANT: Genentech, Inc.  
; APPLICANT: Ashkenazi, Avi  
; APPLICANT: Botstein, David  
; APPLICANT: Desnovers, Luc  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Ferrara, Napoleone  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Fong, Sherman  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerber, Hanspeter  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, A.  
; APPLICANT: Godowski, Paul J.

APPLICANT: Grimaldi, Christopher J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Hillan, Kenneth, J.  
APPLICANT: Kljavin, Ivar J.  
APPLICANT: Mather, Jennie P.  
APPLICANT: Pan, James  
APPLICANT: Paoni, Nicholas F.  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Williams, P. Mickey  
APPLICANT: Wood, William, I.  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
FILE REFERENCE: 10466-14  
CURRENT FILING DATE: 2001-07-18  
PRIOR APPLICATION NUMBER: US/09/909,0888  
PRIOR FILING DATE: 2000-02-22  
PRIOR APPLICATION NUMBER: PCT/US00/04414  
PRIOR FILING DATE: 2000-02-22  
PRIOR APPLICATION NUMBER: US 60/143,048  
PRIOR FILING DATE: 1999-07-07  
PRIOR APPLICATION NUMBER: US 60/145,698  
PRIOR FILING DATE: 1999-07-26  
PRIOR APPLICATION NUMBER: US 60/146,222  
PRIOR FILING DATE: 1999-07-28  
PRIOR APPLICATION NUMBER: PCT/US99/20594  
PRIOR FILING DATE: 1999-09-08  
PRIOR APPLICATION NUMBER: PCT/US99/20944  
PRIOR FILING DATE: 1999-09-13  
PRIOR APPLICATION NUMBER: PCT/US99/21090  
PRIOR FILING DATE: 1999-09-15  
PRIOR APPLICATION NUMBER: PCT/US99/21547  
PRIOR FILING DATE: 1999-09-15  
PRIOR APPLICATION NUMBER: PCT/US99/23089  
PRIOR FILING DATE: 1999-10-05  
PRIOR APPLICATION NUMBER: PCT/US99/28214  
PRIOR FILING DATE: 1999-11-29  
PRIOR APPLICATION NUMBER: PCT/US99/28313  
PRIOR FILING DATE: 1999-11-30  
PRIOR APPLICATION NUMBER: PCT/US99/28564  
PRIOR FILING DATE: 1999-12-02  
PRIOR APPLICATION NUMBER: PCT/US99/28565  
PRIOR FILING DATE: 1999-12-02  
PRIOR APPLICATION NUMBER: PCT/US99/30095  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: PCT/US99/30911  
PRIOR FILING DATE: 1999-12-20  
PRIOR APPLICATION NUMBER: PCT/US99/30999  
PRIOR FILING DATE: 1999-12-20  
PRIOR APPLICATION NUMBER: PCT/US00/00219  
PRIOR FILING DATE: 2000-01-05  
NUMBER OF SEQ ID NOS: 423  
SEQ ID NO 177  
LENGTH: 654  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-909-088B-177

Query Match 70.2%; Score 40; DB 10; Length 654;  
Best Local Similarity 75.0%; Pred. No. 1.le+02;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 LPRWPPPO 8

Db 526 LPKWPYPQ 533  
||:|:|:|

## RESULT 4

US-10-052-586-26  
; Sequence 26, Application US/10052586  
; Patent No. US20020127584A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.

APPLICANT: Chen, Jian  
APPLICANT: Desnoyers, Luc  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Pan, James  
APPLICANT: Smith, Victoria  
APPLICANT: Watanabe, Colin K.  
APPLICANT: Wood, William I.  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P3430R1C1  
CURRENT FILING DATE: 2002-01-15  
PRIOR APPLICATION NUMBER: US/10/052,586  
PRIOR FILING DATE: 1997-09-18  
PRIOR APPLICATION NUMBER: 60/059263  
PRIOR FILING DATE: 1997-09-18  
PRIOR APPLICATION NUMBER: 60/059266  
PRIOR FILING DATE: 1997-09-18  
PRIOR APPLICATION NUMBER: 60/062250  
PRIOR FILING DATE: 1997-10-17  
PRIOR APPLICATION NUMBER: 60/063120  
PRIOR FILING DATE: 1997-10-24  
PRIOR APPLICATION NUMBER: 60/063121  
PRIOR FILING DATE: 1997-10-24  
PRIOR APPLICATION NUMBER: 60/063486  
PRIOR FILING DATE: 1997-10-21  
PRIOR APPLICATION NUMBER: 60/063540  
PRIOR FILING DATE: 1997-10-28  
PRIOR APPLICATION NUMBER: 60/063541  
PRIOR FILING DATE: 1997-10-28  
PRIOR APPLICATION NUMBER: 60/063544  
PRIOR FILING DATE: 1997-10-28  
PRIOR APPLICATION NUMBER: 60/063564  
PRIOR FILING DATE: 1997-10-28  
PRIOR APPLICATION NUMBER: 60/063734  
PRIOR FILING DATE: 1997-10-29  
PRIOR APPLICATION NUMBER: 60/063870  
PRIOR FILING DATE: 1997-10-31  
PRIOR APPLICATION NUMBER: 60/064103  
PRIOR FILING DATE: 1997-10-31  
PRIOR APPLICATION NUMBER: 60/065311  
PRIOR FILING DATE: 1997-11-13  
PRIOR APPLICATION NUMBER: 60/066120  
PRIOR FILING DATE: 1997-11-21  
PRIOR APPLICATION NUMBER: 60/066466  
PRIOR FILING DATE: 1997-11-24  
PRIOR APPLICATION NUMBER: 60/066772  
PRIOR FILING DATE: 1997-11-24  
PRIOR APPLICATION NUMBER: 60/069335  
PRIOR FILING DATE: 1997-12-11  
PRIOR APPLICATION NUMBER: 60/069425  
PRIOR FILING DATE: 1997-12-12  
PRIOR APPLICATION NUMBER: 60/069870  
PRIOR FILING DATE: 1997-12-17  
PRIOR APPLICATION NUMBER: 60/068017  
PRIOR FILING DATE: 1997-12-18  
PRIOR APPLICATION NUMBER: 60/077450  
PRIOR FILING DATE: 1998-03-10  
PRIOR APPLICATION NUMBER: 60/077632  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/077649  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/078886  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/078939  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/079664  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079786  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/080107  
PRIOR FILING DATE: 1998-03-31

;; PRIOR APPLICATION NUMBER: 60/080194  
;; PRIOR FILING DATE: 1998-03-31  
;; PRIOR APPLICATION NUMBER: 60/080327  
;; PRIOR FILING DATE: 1998-04-01  
;; PRIOR APPLICATION NUMBER: 60/080333  
;; PRIOR FILING DATE: 1998-04-01  
;; PRIOR APPLICATION NUMBER: 60/081049  
;; PRIOR FILING DATE: 1998-04-08  
;; PRIOR APPLICATION NUMBER: 60/081070  
;; PRIOR FILING DATE: 1998-04-08  
;; PRIOR APPLICATION NUMBER: 60/081195  
;; PRIOR FILING DATE: 1998-04-09  
;; PRIOR APPLICATION NUMBER: 60/081838  
;; PRIOR FILING DATE: 1998-04-15  
;; PRIOR APPLICATION NUMBER: 60/082568  
;; PRIOR FILING DATE: 1998-04-21  
;; PRIOR APPLICATION NUMBER: 60/082569  
;; PRIOR FILING DATE: 1998-04-21  
;; PRIOR APPLICATION NUMBER: 60/082704  
;; PRIOR FILING DATE: 1998-04-22  
;; PRIOR APPLICATION NUMBER: 60/082797  
;; PRIOR FILING DATE: 1998-04-22  
;; PRIOR APPLICATION NUMBER: 60/083322  
;; PRIOR FILING DATE: 1998-04-28  
;; PRIOR APPLICATION NUMBER: 60/083495  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/083496  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/083499  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/083559  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/084366  
;; PRIOR FILING DATE: 1998-05-05  
;; PRIOR APPLICATION NUMBER: 60/084414  
;; PRIOR FILING DATE: 1998-05-06  
;; PRIOR APPLICATION NUMBER: 60/084639  
;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/084640  
;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/084643  
;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/085573  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085579  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085580  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085582  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085700  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/086023  
;; PRIOR FILING DATE: 1998-05-18  
;; PRIOR APPLICATION NUMBER: 60/086392  
;; PRIOR FILING DATE: 1998-05-22  
;; PRIOR APPLICATION NUMBER: 60/086486  
;; PRIOR FILING DATE: 1998-05-22  
;; PRIOR APPLICATION NUMBER: 60/087098  
;; PRIOR FILING DATE: 1998-05-28  
;; PRIOR APPLICATION NUMBER: 60/087208  
;; PRIOR FILING DATE: 1998-05-28  
;; PRIOR APPLICATION NUMBER: 60/087609  
;; PRIOR FILING DATE: 1998-06-02  
;; PRIOR APPLICATION NUMBER: 60/087759  
;; PRIOR FILING DATE: 1998-06-02  
;; PRIOR APPLICATION NUMBER: 60/087827  
;; PRIOR FILING DATE: 1998-06-03  
;; PRIOR APPLICATION NUMBER: 60/088025  
;; PRIOR FILING DATE: 1998-06-04  
;; PRIOR APPLICATION NUMBER: 60/088028  
;; PRIOR FILING DATE: 1998-06-04  
;; PRIOR APPLICATION NUMBER: 60/088029

;; PRIOR FILING DATE: 1998-06-04  
;; PRIOR APPLICATION NUMBER: 60/088033  
;; PRIOR FILING DATE: 1998-06-04  
;; PRIOR APPLICATION NUMBER: 60/088167  
;; PRIOR FILING DATE: 1998-06-05  
;; PRIOR APPLICATION NUMBER: 60/088202  
;; PRIOR FILING DATE: 1998-06-05  
;; PRIOR APPLICATION NUMBER: 60/088212  
;; PRIOR FILING DATE: 1998-06-05  
;; PRIOR APPLICATION NUMBER: 60/088217  
;; PRIOR FILING DATE: 1998-06-05  
;; PRIOR APPLICATION NUMBER: 60/088326  
;; PRIOR FILING DATE: 1998-06-04  
;; PRIOR APPLICATION NUMBER: 60/088655  
;; PRIOR FILING DATE: 1998-06-09  
;; PRIOR APPLICATION NUMBER: 60/088722  
;; PRIOR FILING DATE: 1998-06-10  
;; PRIOR APPLICATION NUMBER: 60/088738  
;; PRIOR FILING DATE: 1998-06-10  
;; PRIOR APPLICATION NUMBER: 60/088740  
;; PRIOR FILING DATE: 1998-06-10  
;; PRIOR APPLICATION NUMBER: 60/088811  
;; PRIOR FILING DATE: 1998-06-10  
;; PRIOR APPLICATION NUMBER: 60/088824  
;; PRIOR FILING DATE: 1998-06-10  
;; PRIOR APPLICATION NUMBER: 60/088825  
;; PRIOR FILING DATE: 1998-06-10  
;; PRIOR APPLICATION NUMBER: 60/088826  
;; PRIOR FILING DATE: 1998-06-10  
;; PRIOR APPLICATION NUMBER: 60/088861  
;; PRIOR FILING DATE: 1998-06-11  
;; PRIOR APPLICATION NUMBER: 60/088863  
;; PRIOR FILING DATE: 1998-06-11  
;; PRIOR APPLICATION NUMBER: 60/088876  
;; PRIOR FILING DATE: 1998-06-11  
;; PRIOR APPLICATION NUMBER: 60/089090  
;; PRIOR FILING DATE: 1998-06-12  
;; PRIOR APPLICATION NUMBER: 60/089105  
;; PRIOR FILING DATE: 1998-06-12  
;; PRIOR APPLICATION NUMBER: 60/089512  
;; PRIOR FILING DATE: 1998-06-16  
;; PRIOR APPLICATION NUMBER: 60/089514  
;; PRIOR FILING DATE: 1998-06-16  
;; PRIOR APPLICATION NUMBER: 60/089538  
;; PRIOR FILING DATE: 1998-06-17  
;; PRIOR APPLICATION NUMBER: 60/089598  
;; PRIOR FILING DATE: 1998-06-17  
;; PRIOR APPLICATION NUMBER: 60/089653  
;; PRIOR FILING DATE: 1998-06-17  
;; PRIOR APPLICATION NUMBER: 60/089908

Query Match 70.2%; Score 40; DB 12; Length 654;  
Best Local Similarity 75.0%; Pred. No. 1.1e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LPRWPPQ 8  
Db 526 LPKWYPQ 533

RESULT 5

US-09-864-761-44755  
; Sequence 44755, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aecmca-x-1  
; CURRENT APPLICATION NUMBER: US/09/864,761

; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 44755  
; LENGTH: 43  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC019194.2  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.55  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.48  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.49  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.64  
; OTHER INFORMATION: EST HUMAN HIT: AA283201.1, EVALUATE 7.00e+00  
; OTHER INFORMATION: SWISSPROT HIT: O27719, EVALUATE 1.20e+00  
US-09-864-761-44755

Query Match 68.4%; Score 39; DB 10; Length 43;  
Best Local Similarity 85.7%; Pred. No. 12;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LPRWPP 7  
| | | | |  
Db 9 LPRWGP 15

RESULT 6  
US-09-764-868-831  
; Sequence 831, Application US/09764868  
; Patent No. US20020168711A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT232  
; CURRENT APPLICATION NUMBER: US/09/764,868  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - refer to PALM or file wrapper

; NUMBER OF SEQ ID NOS: 1510  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 831  
; LENGTH: 229  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (102)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (125)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (126)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-868-831

Query Match 68.4%; Score 39; DB 9; Length 229;  
Best Local Similarity 75.0%; Pred. No. 56;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LPRWPPQ 8  
| | | | |  
Db 32 LPEWVPQ 39

## RESULT 7

US-09-371-900-40  
; Sequence 40, Application US/09371900  
; Patent No. US20020137700A1  
; GENERAL INFORMATION:  
; APPLICANT: FALB, DEAN A  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE  
; NUMBER OF SEQUENCES: 54  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2711

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/371.900  
FILING DATE: 11-Aug-1999  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/599,654  
FILING DATE: 09-FEB-1996  
APPLICATION NUMBER: US 08/485,573  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/386,844  
FILING DATE: 10-FEB-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: CORUZZI, LAURA A  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7853-104  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 40:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1481 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: unknown

MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 40:  
US-09-371-900-40

Query Match 68.4%; Score 39; DB 10; Length 1481;  
Best Local Similarity 85.7%; Pred. No. 3.3e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 RWPPPOL 9  
DB 8 RWPPPLL 14

RESULT 8

US-09-924-417-60  
Sequence 60, Application US/09924417  
Patent No. US2002014241A1  
GENERAL INFORMATION:  
APPLICANT: Falb, Dean  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
THE TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR  
DISEASE

NUMBER OF SEQUENCES: 67  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036-2711

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq version 2.0

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/924,417  
FILING DATE: 07-Aug-2001  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/034,286  
FILING DATE: 04-MAR-1998  
APPLICATION NUMBER: 08/870,434  
FILING DATE: 06-JUN-1997  
APPLICATION NUMBER: 08/799,910  
FILING DATE: 13-FEB-1997  
APPLICATION NUMBER: 60/011,787  
FILING DATE: 16-FEB-1996  
APPLICATION NUMBER: 08/599,654  
FILING DATE: 09-FEB-1996  
APPLICATION NUMBER: 08/485,573  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: 08/386,844  
FILING DATE: 10-FEB-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7853-114-999

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)7909090  
TELEFAX: (212)8699741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 60:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1481 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 60:

US-09-924-417-60

Query Match 68.4%; Score 39; DB 10; Length 1481;

Best Local Similarity 85.7%; Pred. No. 3.3e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 RWPPPOL 9  
DB 8 RWPPPLL 14

RESULT 9

US-09-864-761-42815  
Sequence 42815, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aecomica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/006666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/006667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/006664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/006669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/006665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/006668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/006663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/006662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/006661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/006670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117

SOFTWARE: Anomax Sequence Listing Engine vers. 1.1

SEQ ID NO 42815

LENGTH: 71

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

OTHER INFORMATION: MAP TO AC009946.1  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.4  
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.98  
OTHER INFORMATION: EST HUMAN HIT: AA873561.1, EVALUE 2.00e-20  
OTHER INFORMATION: SWISSPROT HIT: P10069, EVALUE 1.90e+00

US-09-864-761-42815

Query Match 66.7%; Score 38; DB 10; Length 71;  
Best Local Similarity 83.3%; Pred. No. 25;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 PRWPPP 7  
| | | | |  
Db 40 PAWPPP 45

## RESULT 10

US-09-764-898-188  
; Sequence 188, Application US/09764898  
; Patent No. US20020090673A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: P0201

; CURRENT APPLICATION NUMBER: US/09/764, 898

; CURRENT FILING DATE: 2001-01-17

; Prior application data removed - consult PALM or file wrapper

; NUMBER OF SEQ ID NOS: 311

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 188

; LENGTH: 348

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-764-898-188

Query Match 66.7%; Score 38; DB 10; Length 348;  
Best Local Similarity 83.3%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 PRWPPP 7  
| | | | |  
Db 142 PTWPPP 147

## RESULT 11

US-09-764-853-568

; Sequence 568, Application US/09764853

; Patent No. US20020090672A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: P0206

; CURRENT APPLICATION NUMBER: US/09/764, 853

; CURRENT FILING DATE: 2001-01-17

; Prior application data removed - consult PALM or file wrapper

; NUMBER OF SEQ ID NOS: 939

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 568

; LENGTH: 482

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: SITE

; LOCATION: (188)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

; NAME/KEY: SITE

; LOCATION: (194)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

; NAME/KEY: SITE

; LOCATION: (215)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

; NAME/KEY: SITE

; LOCATION: (281)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-764-853-568

Query Match 66.7%; Score 38; DB 10; Length 482;  
Best Local Similarity 83.3%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 PRWPPP 7  
| | | | |  
Db 251 PTWPPP 256

## RESULT 12

US-09-866-582-36

; Sequence 36, Application US/09866582

; Patent No. US20020127620A1

; GENERAL INFORMATION:

; APPLICANT: Witman, George B.

; APPLICANT: Pazour, Gregory J.

; APPLICANT: Rosenbaum, Joel L.

; APPLICANT: Cole, Douglas G.

; TITLE OF INVENTION: INTRAFAGELLAR TRANSPORT

; FILE REFERENCE: 07917-145001

; CURRENT APPLICATION NUMBER: US/09/866, 582

; CURRENT FILING DATE: 2001-05-24

; PRIOR APPLICATION NUMBER: US 60/206, 923

; PRIOR FILING DATE: 2000-05-24

; NUMBER OF SEQ ID NOS: 50

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 36

; LENGTH: 1198

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-866-582-36

Query Match 66.7%; Score 38; DB 10; Length 1198;  
Best Local Similarity 71.4%; Pred. No. 3.7e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LPRWPPP 7  
: | | | | |  
Db 1140 IKRWPPP 1146

## RESULT 13

US-10-001-215-5

; Sequence 5, Application US/10001215

; Patent No. US20020147323A1

; GENERAL INFORMATION:

; APPLICANT: Bandaru Rajasekhkar

; APPLICANT: Rosana Kapeller-Libermann

; TITLE OF INVENTION: 16224 and 69611, NOVEL HUMAN KINASES AND

; TITLE OF INVENTION: USES THEREOF

; FILE REFERENCE: MNI-209

; CURRENT APPLICATION NUMBER: US/10/001,215

; CURRENT FILING DATE: 2001-11-30

; PRIOR APPLICATION NUMBER: 60/250,917

; PRIOR FILING DATE: 2000-11-03

; NUMBER OF SEQ ID NOS: 6

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 5

; LENGTH: 1241

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-001-215-5

Query Match 66.7%; Score 38; DB 12; Length 1241;  
Best Local Similarity 71.4%; Pred. No. 3.8e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LPRWPPP 7  
: | | | | |  
Db 1183 IKRWPPP 1189

## RESULT 14

US-09-864-761-46140

; Sequence 46140, Application US/09864761

; Patent No. US20020048763A1

; GENERAL INFORMATION:

APPLICANT: Penn, Sharron G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aeomica-X-1  
CURRENT APPLICATION NUMBER: US 09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 46140  
LENGTH: 39  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC023078.2  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.1  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7  
OTHER INFORMATION: EST HUMAN HIT: BE734169.1, EVALUE 9.00e-09  
OTHER INFORMATION: SWISSPROT HIT: P31695, EVALUE 3.70e+00

US-09-864-761-46140  
Query Match 64.9%; Score 37; DB 10; Length 39;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RWPPP 7  
|||||  
Db 22 RWPPP 26

RESULT 15  
US-09-864-761-40064  
Sequence 40064, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharron G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO  
FILE REFERENCE: Aeomica-X-1  
CURRENT APPLICATION NUMBER: US 09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 40064  
LENGTH: 52  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC004408.1  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.5  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.7  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.5  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.4  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.1  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.8  
OTHER INFORMATION: EST HUMAN HIT: AW952275.1, EVALUE 2.00e-12  
OTHER INFORMATION: SWISSPROT HIT: P03204, EVALUE 2.70e+00

US-09-864-761-40064  
Query Match 64.9%; Score 37; DB 10; Length 52;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



Qy 2 PRWPP 6  
IIII  
Db 42 PRWPP 46

Search completed: November 26, 2002, 19:44:26  
Job time : 4.74194 secs

**THIS PAGE BLANK (SPT0)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 25, 2002, 17:00:51 ; Search time 65 seconds  
(without alignments)  
8188.147 Million cell updates/sec

Title: US-09-674-593-1

Perfect score: 1382

Sequence: 1 cattatgctaacagcataaa.....taagcaattaagtctctg 1382

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 341543 seqs, 192557720 residues

Total number of hits satisfying chosen parameters: 683086

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications\_NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	48.4	3.5	2000	9	US-09-938-842A-3267
c 2	48.4	3.5	2004	10	Sequence 3267, Ap
c 3	48.4	3.5	2004	10	Sequence 211, App
c 4	45.6	3.3	257	10	Sequence 297, App
c 5	44.6	3.2	1967	10	Sequence 14300, A
c 6	43.4	3.1	2000	9	Sequence 565, App
c 7	42.8	3.1	239	10	Sequence 4451, Ap
c 8	42.6	3.1	364	10	Sequence 11438, A
c 9	42.2	3.1	3238	10	Sequence 10067, A
c 10	41.8	3.0	2136	9	Sequence 5, Appli
c 11	41.8	3.0	2136	10	Sequence 302, App
c 12	41.8	3.0	2136	10	Sequence 302, App
c 13	41.8	3.0	2136	10	Sequence 302, App
c 14	41.8	3.0	2136	10	Sequence 302, App
c 15	41.8	3.0	2136	10	Sequence 302, App
c 16	41.8	3.0	2136	10	Sequence 302, App
c 17	41.8	3.0	2136	10	Sequence 302, App
c 18	41.8	3.0	2136	10	Sequence 302, App
c 19	41.8	3.0	2136	10	Sequence 302, App

#### ALIGNMENTS

##### RESULT 1

US-09-938-842A-3267/c  
; Sequence 3267, Application US/09938842A

; Patent No. US20020160378A1

; GENERAL INFORMATION:

; APPLICANT: Harper, Jeff

; APPLICANT: Krieps, Joel

; APPLICANT: Wang, Xun

; APPLICANT: Zhu, Tong

; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING

; FILE REFERENCE: SAME, AND METHODS OF USE

; CURRENT APPLICATION NUMBER: US/09/938.842A

; CURRENT FILING DATE: 2001-08-24

; PRIOR APPLICATION NUMBER: US 60/227,866

; PRIOR FILING DATE: 2000-08-24

; PRIOR APPLICATION NUMBER: US 60/264,647

; PRIOR FILING DATE: 2001-01-16

; PRIOR APPLICATION NUMBER: US 60/300,111

; PRIOR FILING DATE: 2001-06-22

; NUMBER OF SEQ ID NOS: 5379

; SEQ ID NO 3267

; LENGTH: 2000

; TYPE: DNA

; ORGANISM: Arabidopsis thaliana

US-09-938-842A-3267

Query Match 3.5%; Score 48.4; DB 9; Length 2000;

Best Local Similarity 57.7%; Pred. No. 0.03;

Matches 105; Conservative 0; Mismatches 76; Indels 1; Gaps 1;

Qy	89	ATAATTTAAACAGPACGGAAATCGAGCAATTTCTCTCTCTGTCACATATAGTGC	148
Db	1962	AGAAATATAAGAGGAAGTATCTTCTTCTTCAACGATGATGATGATGATTT	1903
Qy	149	CAACACTTGGAAAGTATTTTAAGAATCTTTATTAATAAAGAGATGGA-TTTC	207
Db	1902	GAAATCTCAGAGACCTTCTTTCTTTATCTATATAACATGAGGAATTTT	1843
Qy	208	GAAAAAATAAGGAAGGAAGAAAGAAAGAAAGAAAGAAAGAAAGTATCAGTTG	267
Db	1842	AAAAAATAATTTTAAAAAAGAAAAAGTAAAAAAGTAAAAAAGTAAAAAATTT	1783

Sequence 302, App  
Sequence 302, App  
Sequence 302, App  
Sequence 227, App  
Sequence 9576, Ap  
Sequence 10182, A  
Sequence 11234, A  
Sequence 315, App  
Sequence 351, App  
Sequence 13, Appl  
Sequence 5093, Ap  
Sequence 296, App  
Sequence 1, Appli  
Sequence 4582, Ap  
Sequence 393, App  
Sequence 1, Appli  
Sequence 4438, Ap  
Sequence 25, Appl  
Sequence 4161, Ap  
Sequence 9335, Ap  
Sequence 49, Appl  
Sequence 31, Appl  
Sequence 8148, Ap  
Sequence 335, App

	Query Match Best Local Matches	3.3%; Similarity 99;	Score 52.7%; Conservative	DB 10; 0.053; 89;	Length 257; 0; Gaps
Qy	209	AAAAAAAAATTAAGGAAAGGAAAGAAAAAACTGAACACAGAAAAACGCAAAAGTATCATGCTTTGG	268		
Db	254	AAAAAAAAATAATAATAATCATATTAA	195		
Qy	269	TCACATAACCTTTGCAAGGATACCTTTTTATTCTTTTAAAGATTCTCTGTTGTTTATACACA	328		
Db	194	TATCAATAATCTTAATATATAATAATAATAATAATCTTTATATACCTTTTAAATATGAATAA	135		
Qy	329	GATTTTAAGTTTACTCTCTACTGCTGACCCCAAGTCAAAATTCCTCTCCACAGTCACAGTGCA	388		
Db	134	TATTATTAGTATATATCATCAAGCATATTCGAAGTTCATTTTATTATGCAATTCATATAGTCT	75		

Qy 389 ACCTCTAC 396  
Db 74 ATATCTAC 67

## RESULT 5

```

US-09-925-301-565
; Sequence 365, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 565
; LENGTH: 1967
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-301-565

```

Query Match	3.2%	Score 44.6:	DB 10:	Length 1967:
Best Local Similarity	53.1%	Pred. No. 0.25:		
Matches	95;	Conservative	0;	Mismatches 84;
				Indels 0;
				Gaps 0;
Qy 71	TTGGAATCAAAGGTTGAAATAATTTAAACACGACGCGGGAATTCACAGAGCAATTTTCTCCCTC	130		
Db 1777	TGGCAGTATATATTAGTGTGATAATTTGTATTTTTCACAAAAAAAAGTTTAAACTCTCTCTT	1836		
Qy 131	TGGTGACAATATAGTGCCAACACCTTGGAACTGATTTTTTAAGATGTGTTATTTAAATTTAA	190		
Db 1837	TCTTTTTATTATATGACCAGCTTTTGGTATTTTCATTCTTACCAAGTCTATTTTTAGAA	1896		
Qy 191	AAGGATGGATTTCCAAGGAAAAAAAATAAGGAAAGGAAAGAAAAAACTGAACAGAAAA	249		
Db 1897	TAAATTTGTTCTCTCTCTCTTCAAAAAAAAATAAAAAAAAATAAAAAAAAATAAAAAAAA	1955		

RESULT 6  
 US-09-938-842A-4451  
 : Sequence 4451, Application US/09938842A  
 : Patent No. US20020160378A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Harper, Jeff  
 : APPLICANT: Kreps, Joel  
 : APPLICANT: Wang, Xun  
 : APPLICANT: Zhu, Tong  
 : TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
 : TITLE OF INVENTION: SAME, AND METHODS OF USE  
 : FILE REFERENCE: SCRIPI300-3  
 : CURRENT APPLICATION NUMBER: US/09/938, 842A  
 : CURRENT FILING DATE: 2001-08-24  
 : PRIOR APPLICATION NUMBER: US 60/227,866  
 : PRIOR FILING DATE: 2000-08-24  
 : PRIOR APPLICATION NUMBER: US 60/264,647  
 : PRIOR FILING DATE: 2001-01-16  
 : PRIOR APPLICATION NUMBER: US 60/300,111  
 : PRIOR FILING DATE: 2001-06-22  
 : NUMBER OF SEQ ID NOS: 5379  
 : SEQ ID NO 4451  
 : LENGTH: 2000  
 : TYPE: DNA  
 : ORGANISM: Arabidopsis thaliana  
 US-09-938-842A-4451

Query Match 3.1%; Score 43.4; DB 9; Length 2000;  
Best Local Similarity 51.3%; Pred. No. 0.5;

[illegible]

## RESULT 7

```

RESULT: /
US-09-960-352-11438/c
; Sequence 11438, Application US/09960352
; Patent NO. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES A
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 11438
; LENGTH: 239
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 49-LIB3058-019-Q1-K1-E1C
US-09-960-352-11438

```

	Query Match	3.1%;	Score 42.8;	DB 10;	Length 239;
	Best Local Similarity	52.2%;	Pred. No. 0.25;		
Matches	95; Conservative	0;	Mismatches 87;	Indels 0;	Gaps 0;
Qy	75	AATGAAGGTTGAAATAATTTAAACAGFACGGGAAATGCAGAGCAATTTTCTCCTCTGGT	134		
Db	235	AATTAAATAGNATATTTATTACTTTTAAGCCAAAGTGAGGCCIAATTAGTTTTTTTT	176		
Qy	i35	GACAATATAGTGTCCAACACITGGAGTGATTTTAAAGAATGTTTATTTAAATATAAAGG	194		
Db	175	TATATAGAATATTTCTACATTTTAAATTTCTTTTAAAAAAAACAAAAAAAAAAAAA	116		
Qy	195	ATGGATTTCCAGGAAAAAATAATAGGAAAAAGGAAAAAACCTGAACAGAAACGCCAA	254		
Db	115	AAACCA	56		
Qy	255	AA	256		
Db	55	AA	54		

RESULTS

RESULT B  
US-09-960-352-10067/c  
: Sequence 10067, Application US/09960352  
: Patent No. US20020137139A1  
: GENERAL INFORMATION:  
: APPLICANT: Warren, Wesley C.  
: APPLICANT: Tao, Nengbing  
: APPLICANT: Byatt, John C.  
: APPLICANT: Mathialagan, Nagappan  
: TITLE OF INVENTION: NUCLEIC ACID AND  
: TITLE OF INVENTION: MUSCLE AND FAT O

FILE REFERENCE: 16511.006/37-21(10298)/C  
CURRENT APPLICATION NUMBER: US/09/960,352  
CURRENT FILING DATE: 2001-09-24  
NUMBER OF SEQ ID NOS: 15112  
SEQ ID NO 10067  
LENGTH: 364  
TYPE: DNA  
ORGANISM: Bos taurus  
OTHER INFORMATION: Clone ID: 43-LIB34-031-Q1-E1-C4  
US-09-960-352-10067

Query Match 3.1% Score 42.6; DB 10; Length 364;  
Best Local Similarity 49.3%; Pred. No. 0.34;  
Matches 111; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

QY 35 GAGCAGGTACAAAAGTGGTGTGCAATTCCTGGAATGAAGGTTGAATATT 94  
DB 312 GACCAGGGGATAATGGCTGCTGGATTCAGATGATCTGTAGGAATCTTTGGCCATCAA 253  
QY 95 TAAACAGTACGGGAAATGCAGACAAATTTCTCTCTGTCGACAAATATAGTGTCCACAC 154  
DB 252 CATCAGTTTCTGATTAACACAGTCTCTGTGTTCTCTGTCGAAAGTAACAGTGTCCCAT 193  
QY 155 TTGGAAGTGAATTTAAGAAATTTTATTAATTAAGAGTGGATTTCCAGGAAAAA 214  
DB 192 GCGGTGAATTTGTTGTTAGTTAGTTGTTGATTAACACTTTTGTGTTTCTCAAAAAA 133  
QY 215 RATAAGGAAAGGAAAGAAAAACTGACAGACAAACGCAAAAGTA 259  
DB 132 AAAAAAGAAAAATAAAAAAGAAAAAAGAAAAAAGAAAAAAGAA 88

RESULT 9  
US-09-874-628-5  
Sequence 5, Application US/09874628  
Patent No. US20020137133A1  
GENERAL INFORMATION:  
APPLICANT: WOZNEY, John  
CELESTE, Anthony J.  
THIES, R. Scott  
YAMAJI, No. US20020137133A1  
TITLE OF INVENTION: RECEPTOR PROTEINS  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute Inc.- Legal Affairs  
STREET: 87 CambridgePark Drive  
CITY: Cambridge  
STATE: MA  
COUNTRY: USA  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/874,628  
FILING DATE: 05-Jun-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/123,934  
FILING DATE: 17-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: LAZAR, Steven R.  
REGISTRATION NUMBER: 32,618  
REFERENCE/DOCKET NUMBER: 5203  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617 876 1170  
TELEFAX: 617 876 5851  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3238 base pairs  
TYPE: nucleic acid

STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
IMMEDIATE SOURCE:  
CLONE: CFK1-10a  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 474..2000  
SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
US-09-874-628-5

Query Match 3.1% Score 42.2; DB 10; Length 3238;  
Best Local Similarity 50.8%; Pred. No. 1.2;  
Matches 101; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 51 GTGAGTGTGTGCAATTCCTGGAATGAAAGGTTGAAATAATTTAAACAGTACGGGAAA 110  
DB 3021 GGGGGGAAATGCATCTCTTCGGAACATATCCATTACATGCTTTAACTCTGCCAGAAA 3080  
QY 111 TGCAGAGCAATTTCTCTCTGCTGGTGACAATATATAGTGTCCACACTTGGAAAGTATT 170  
DB 3081 AAAAAATAACTATTTTGTGTTTAAATCTACTTTTGTATTTAGTAGTTATTTCTATAAATTA 3140  
QY 171 AGAATGTTTATTTAAATTAAGGATGGATTTCCAGGAAAAAATAAGAAAAAGGAAA 230  
DB 3141 AATAACTGTTTTCAGTCAAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 3200  
QY 231 GAAAAAACTGAACAGAAAA 249  
DB 3201 AAAAAAATAAAAAA 3219

RESULT 10  
US-09-992-598-302  
Sequence 302, Application US/09992598  
Patent No. US20020160384A1  
GENERAL INFORMATION:  
APPLICANT: Ashkenazi, Avi J.  
APPLICANT: Baker, Kevin P.  
APPLICANT: Botstein, David  
APPLICANT: Desnoyers, Luc  
APPLICANT: Eaton, Dan L.  
APPLICANT: Ferrara, Napoleone  
APPLICANT: Fong, Sherman  
APPLICANT: Gerber, Hanspeter  
APPLICANT: Gerritsen, Mary E.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Grimaldi, J. Christopher  
APPLICANT: Gurney, Austin L.  
APPLICANT: Kljavin, Ivar J.  
APPLICANT: Napier, Mary A.  
APPLICANT: Pan, James  
APPLICANT: Paoni, Nicholas F.  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tamas, Daniel  
APPLICANT: Tamas, Colin K.  
APPLICANT: Williams, P. Mickey  
APPLICANT: Wood, William I.  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
FILE REFERENCE: P2730PIC20  
CURRENT APPLICATION NUMBER: US/09/992,598  
CURRENT FILING DATE: 2001-11-14  
PRIOR APPLICATION NUMBER: 60/049787  
PRIOR FILING DATE: 1997-06-16  
PRIOR APPLICATION NUMBER: 60/062250  
PRIOR FILING DATE: 1997-10-17  
PRIOR APPLICATION NUMBER: 60/065186  
PRIOR FILING DATE: 1997-11-12  
PRIOR APPLICATION NUMBER: 60/065311

PRIOR APPLICATION NUMBER:	60/089532
PRIOR FILING DATE:	1998-06-17
PRIOR APPLICATION NUMBER:	60/089538
PRIOR FILING DATE:	1998-06-17
PRIOR APPLICATION NUMBER:	60/089598
PRIOR FILING DATE:	1998-06-17
PRIOR APPLICATION NUMBER:	60/089599
PRIOR FILING DATE:	1998-06-17
PRIOR APPLICATION NUMBER:	60/089600
PRIOR FILING DATE:	1998-06-17
PRIOR APPLICATION NUMBER:	60/089653
PRIOR FILING DATE:	1998-06-17
PRIOR APPLICATION NUMBER:	60/089801
PRIOR FILING DATE:	1998-06-18
PRIOR APPLICATION NUMBER:	60/089907
PRIOR FILING DATE:	1998-06-18
PRIOR APPLICATION NUMBER:	60/089908
PRIOR FILING DATE:	1998-06-18
PRIOR APPLICATION NUMBER:	60/089947
PRIOR FILING DATE:	1998-06-19
PRIOR APPLICATION NUMBER:	60/089948
PRIOR FILING DATE:	1998-06-19
PRIOR APPLICATION NUMBER:	60/089952
PRIOR FILING DATE:	1998-06-19
PRIOR APPLICATION NUMBER:	60/090246
PRIOR FILING DATE:	1998-06-22
PRIOR APPLICATION NUMBER:	60/090252
PRIOR FILING DATE:	1998-06-22
PRIOR APPLICATION NUMBER:	60/090254
PRIOR FILING DATE:	1998-06-22
PRIOR APPLICATION NUMBER:	60/090349
PRIOR FILING DATE:	1998-06-23
PRIOR APPLICATION NUMBER:	60/090355
PRIOR FILING DATE:	1998-06-23
PRIOR APPLICATION NUMBER:	60/090429
PRIOR FILING DATE:	1998-06-24
PRIOR APPLICATION NUMBER:	60/090431
PRIOR FILING DATE:	1998-06-24
PRIOR APPLICATION NUMBER:	60/090435
PRIOR FILING DATE:	1998-06-24
PRIOR APPLICATION NUMBER:	60/090444
PRIOR FILING DATE:	1998-06-24
PRIOR APPLICATION NUMBER:	60/090445
PRIOR FILING DATE:	1998-06-24
PRIOR APPLICATION NUMBER:	60/090472
PRIOR FILING DATE:	1998-06-24
PRIOR APPLICATION NUMBER:	60/090535
PRIOR FILING DATE:	1998-06-24
PRIOR APPLICATION NUMBER:	60/090540
PRIOR FILING DATE:	1998-06-24
PRIOR APPLICATION NUMBER:	60/090542
PRIOR FILING DATE:	1998-06-24
PRIOR APPLICATION NUMBER:	60/090557
PRIOR FILING DATE:	1998-06-24
PRIOR APPLICATION NUMBER:	60/090676
PRIOR FILING DATE:	1998-06-25
PRIOR APPLICATION NUMBER:	60/090678
PRIOR FILING DATE:	1998-06-25
PRIOR APPLICATION NUMBER:	60/090690
PRIOR FILING DATE:	1998-06-25
PRIOR APPLICATION NUMBER:	60/090696
PRIOR FILING DATE:	1998-06-25
PRIOR APPLICATION NUMBER:	60/090862
PRIOR FILING DATE:	1998-06-26
PRIOR APPLICATION NUMBER:	60/090863
PRIOR FILING DATE:	1998-06-26
PRIOR APPLICATION NUMBER:	60/091360
PRIOR FILING DATE:	1998-07-01
PRIOR APPLICATION NUMBER:	60/091478

```
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091544
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091626
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

Query Match          3.08; Score 41.8; DB 9; Length 2136;
Best Local Similarity 56.08; Pred. No. 1.3;
Matches 79; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 116 AGCAATTTCTCCTCTGCTGACATATAGTGTCCAACACACTTGGAGTGATTTTAAAGAAAT 175
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1979 AACGTATCATTTCTGCTGAGGTGGAGTGTCCCATCTTTTATCAAGGTGATTGTGA 2038
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 176 GTTTATTTAAATTAAGAGTGATTTTCCAGGAAAAAATAAGGAAAAAGAAAAA 235
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2039 TTTTGACTAATAAAGAAATTTGTAAAAAATAAAGAAAAAATAAAGAAAAA 2098
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 236 AACGACAGAAAAAGCAAAA 256
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2099 AAAAAAATAAAGAAAAA 2119
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
US-09-989-722-302
; Sequence 302. Application US/09989722
; Patent No. US20020072067A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P27301C63
; CURRENT APPLICATION NUMBER: US/09/989,722
; PRIOR FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: 60/049787
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12

; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/075945
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/087106
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087759
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087827
; PRIOR FILING DATE: 1998-06-03
; PRIOR APPLICATION NUMBER: 60/088021
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088025
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088026
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088028
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088029
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088030
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088033
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088326
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088167
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088202
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088212
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088217
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088655
; PRIOR FILING DATE: 1998-06-09
; PRIOR APPLICATION NUMBER: 60/088734
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088738
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088742
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088810
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088824
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088826
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088858
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088861
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088876
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/089105
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089440
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089512
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089514
```



; PRIOR APPLICATION NUMBER: 6

Wed Nov 27 08:47:40 2002

[illegible]

```

; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091478
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091544
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091626
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

Query Match      3.0%  Score 41.8; DB 10; Length 2136;
Best Local Similarity 56.0%  Pred. No. 1.3;
Matches 79; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 116 AGCAATTTCTCTCTGTCACATATAGTGTCCACACTTGGAGTGTATTTTAAAGAT 175
Db 1979 AACTGTATCATTTCTGCTGAGGCTGAGTGTCCCATCTCTTTTAACTCAAGGTGATTGGA 2038

Qy 176 GTTTATTAAATTAAGAGTGGATTTCACAGGAAAAAATAAGGAAAGGAAAGAAAA 235
Db 2039 TTTTGACTAATAAAAAAGATTTCTTAAAAAATAAAAAAATAAAAAAATAAAAAA 2098

Qy 236 AACTGAACAGAAACGCAAAA 256
Db 2099 AAAAAAATAAAAAAATAAAAA 2119

RESULT 13
US-09-989-279-302
; Sequence 302, Application US/09989279
; Patent No. US20020072496A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Kijavlin, Ivar J.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas P.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2730P1C56
; CURRENT APPLICATION NUMBER: US/09/989,279
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: 60/049787
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/075945
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/087106
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087759
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087827
; PRIOR FILING DATE: 1998-06-03
; PRIOR APPLICATION NUMBER: 60/088021
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088025
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088026
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088028
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088029
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088030
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088033
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088326
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088167
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088202
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088212
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088217
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088655
; PRIOR FILING DATE: 1998-06-09
; PRIOR APPLICATION NUMBER: 60/088734
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088738
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088742
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088810
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088824
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088826
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088858
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088861
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088876
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/089105
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089440
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089512
```

;  
; PRIOR FILING DATE: 1998-06-16  
; PRIOR APPLICATION NUMBER: 60/089514  
; PRIOR FILING DATE: 1998-06-16  
; PRIOR APPLICATION NUMBER: 60/089532  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089538  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089598  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089599  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089600  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089653  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089801  
; PRIOR FILING DATE: 1998-06-18  
; PRIOR APPLICATION NUMBER: 60/089907  
; PRIOR FILING DATE: 1998-06-18  
; PRIOR APPLICATION NUMBER: 60/089908  
; PRIOR FILING DATE: 1998-06-18  
; PRIOR APPLICATION NUMBER: 60/089947  
; PRIOR FILING DATE: 1998-06-19  
; PRIOR APPLICATION NUMBER: 60/089948  
; PRIOR FILING DATE: 1998-06-19  
; PRIOR APPLICATION NUMBER: 60/089952  
; PRIOR FILING DATE: 1998-06-19  
; PRIOR APPLICATION NUMBER: 60/090246  
; PRIOR FILING DATE: 1998-06-22  
; PRIOR APPLICATION NUMBER: 60/090252  
; PRIOR FILING DATE: 1998-06-22  
; PRIOR APPLICATION NUMBER: 60/090254  
; PRIOR FILING DATE: 1998-06-22  
; PRIOR APPLICATION NUMBER: 60/090349  
; PRIOR FILING DATE: 1998-06-23  
; PRIOR APPLICATION NUMBER: 60/090355  
; PRIOR FILING DATE: 1998-06-23  
; PRIOR APPLICATION NUMBER: 60/090429  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 60/090431  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 60/090445  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 60/090472  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 60/090535  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 60/090540  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 60/090542  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 60/090557  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 60/090676  
; PRIOR FILING DATE: 1998-06-25  
; PRIOR APPLICATION NUMBER: 60/090678  
; PRIOR FILING DATE: 1998-06-25  
; PRIOR APPLICATION NUMBER: 60/090690  
; PRIOR FILING DATE: 1998-06-25  
; PRIOR APPLICATION NUMBER: 60/090694  
; PRIOR FILING DATE: 1998-06-25  
; PRIOR APPLICATION NUMBER: 60/090695  
; PRIOR FILING DATE: 1998-06-25  
; PRIOR APPLICATION NUMBER: 60/090696  
; PRIOR FILING DATE: 1998-06-25  
; PRIOR APPLICATION NUMBER: 60/090862  
; PRIOR FILING DATE: 1998-06-26  
; PRIOR APPLICATION NUMBER: 60/090863  
; PRIOR FILING DATE: 1998-06-26  
; PRIOR APPLICATION NUMBER: 60/090863  
; PRIOR FILING DATE: 1998-06-26

;  
; PRIOR APPLICATION NUMBER: 60/091360  
; PRIOR FILING DATE: 1998-07-01  
; PRIOR APPLICATION NUMBER: 60/091478  
; PRIOR FILING DATE: 1998-07-02  
; PRIOR APPLICATION NUMBER: 60/091544  
; PRIOR FILING DATE: 1998-07-01  
; PRIOR APPLICATION NUMBER: 60/091519  
; PRIOR FILING DATE: 1998-07-02  
; PRIOR APPLICATION NUMBER: 60/091626  
; PRIOR FILING DATE: 1998-07-02  
; PRIOR APPLICATION NUMBER: 60/091633  
; PRIOR FILING DATE: 1998-07-02  
; PRIOR APPLICATION NUMBER: 60/091978  
; PRIOR FILING DATE: 1998-07-07  
; PRIOR APPLICATION NUMBER: 60/091982  
; PRIOR FILING DATE: 1998-07-07  
; PRIOR APPLICATION NUMBER: 60/092182  
; PRIOR FILING DATE: 1998-07-09  
;  
Query Match 3.0%; Score 41.8; DB 10; Length 2136;  
Best Local Similarity 56.0%; Pred. No. 1.3; Indels 0; Gaps 0;  
Matches 79; Conservative 0; Mismatches 62  
Qy 116 AGCAATTTCTCTCTGTGACAAATATAGTGTCCCAACACTTGGAGTGTGATTTTAAAGAT 175  
Db 179 AACTGTATCATTTTCTGCTGAGGGTGGAGTGTCCCATCTTTTAAATCAAGGTGATTTGA 2038  
Qy 176 GTTTATTAAATTAAGGATGGATTTCCAGGAAAAAATAAGGAAAGGAAAGAAAA 235  
Db 2039 TTTTGACTAATAAAAAACAATTTGTAAAAAATAAAAAAATAAAAAAATAAAAAA 2098  
Qy 236 AACTGACAGAAACGCAAAA 256  
Db 2099 AAAAAAATAAAAAAATAAAAAA 2119

RESULT 14

US-09-989-727-302  
; Sequence 302, Application US/09989727  
; Patent No. US20020072497A1  
; GENERAL INFORMATION:  
; APPLICANT: Ashkenazi, Avi J.  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Botstein, David  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Ferrara, Napoleone  
; APPLICANT: Fong, Sherman  
; APPLICANT: Gerber, Hanspeter  
; APPLICANT: Gerriksen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, J. Christopher  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Kljavin, Ivar J.  
; APPLICANT: Napier, Mary A.  
; APPLICANT: Pan, James  
; APPLICANT: Paoni, Nicholas F.  
; APPLICANT: Roy, Margaret Ann  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Williams, P. Mickey  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
; TITLE OF INVENTION: Acids Encoding the Same  
; FILE REFERENCE: P2730PLC65  
; CURRENT APPLICATION NUMBER: US/09/989,727  
; CURRENT FILING DATE: 2001-11-19  
; PRIOR APPLICATION NUMBER: 60/049787  
; PRIOR FILING DATE: 1997-06-16  
; PRIOR APPLICATION NUMBER: 60/062250

1	PRIOR APPLICATION NUMBER: 60/089512
2	PRIOR FILING DATE: 1998-06-16
3	PRIOR APPLICATION NUMBER: 60/089514
4	PRIOR FILING DATE: 1998-06-16
5	PRIOR APPLICATION NUMBER: 60/089532
6	PRIOR FILING DATE: 1998-06-17
7	PRIOR APPLICATION NUMBER: 60/089538
8	PRIOR FILING DATE: 1998-06-17
9	PRIOR APPLICATION NUMBER: 60/089598
10	PRIOR FILING DATE: 1998-06-17
11	PRIOR APPLICATION NUMBER: 60/089599
12	PRIOR FILING DATE: 1998-06-17
13	PRIOR APPLICATION NUMBER: 60/089600
14	PRIOR FILING DATE: 1998-06-17
15	PRIOR APPLICATION NUMBER: 60/089653
16	PRIOR FILING DATE: 1998-06-17
17	PRIOR APPLICATION NUMBER: 60/089801
18	PRIOR FILING DATE: 1998-06-18
19	PRIOR APPLICATION NUMBER: 60/089907
20	PRIOR FILING DATE: 1998-06-18
21	PRIOR APPLICATION NUMBER: 60/089908
22	PRIOR FILING DATE: 1998-06-18
23	PRIOR APPLICATION NUMBER: 60/089947
24	PRIOR FILING DATE: 1998-06-19
25	PRIOR APPLICATION NUMBER: 60/089948
26	PRIOR FILING DATE: 1998-06-19
27	PRIOR APPLICATION NUMBER: 60/089952
28	PRIOR FILING DATE: 1998-06-19
29	PRIOR APPLICATION NUMBER: 60/090246
30	PRIOR FILING DATE: 1998-06-22
31	PRIOR APPLICATION NUMBER: 60/090252
32	PRIOR FILING DATE: 1998-06-22
33	PRIOR APPLICATION NUMBER: 60/090254
34	PRIOR FILING DATE: 1998-06-22
35	PRIOR APPLICATION NUMBER: 60/090349
36	PRIOR FILING DATE: 1998-06-23
37	PRIOR APPLICATION NUMBER: 60/090355
38	PRIOR FILING DATE: 1998-06-23
39	PRIOR APPLICATION NUMBER: 60/090429
40	PRIOR FILING DATE: 1998-06-24
41	PRIOR APPLICATION NUMBER: 60/090431
42	PRIOR FILING DATE: 1998-06-24
43	PRIOR APPLICATION NUMBER: 60/090435
44	PRIOR FILING DATE: 1998-06-24
45	PRIOR APPLICATION NUMBER: 60/090444
46	PRIOR FILING DATE: 1998-06-24
47	PRIOR APPLICATION NUMBER: 60/090445
48	PRIOR FILING DATE: 1998-06-24
49	PRIOR APPLICATION NUMBER: 60/090472
50	PRIOR FILING DATE: 1998-06-24
51	PRIOR APPLICATION NUMBER: 60/090535
52	PRIOR FILING DATE: 1998-06-24
53	PRIOR APPLICATION NUMBER: 60/090540
54	PRIOR FILING DATE: 1998-06-24
55	PRIOR APPLICATION NUMBER: 60/090542
56	PRIOR FILING DATE: 1998-06-24
57	PRIOR APPLICATION NUMBER: 60/090557
58	PRIOR FILING DATE: 1998-06-24
59	PRIOR APPLICATION NUMBER: 60/090676
60	PRIOR FILING DATE: 1998-06-25
61	PRIOR APPLICATION NUMBER: 60/090678
62	PRIOR FILING DATE: 1998-06-25
63	PRIOR APPLICATION NUMBER: 60/090690
64	PRIOR FILING DATE: 1998-06-25
65	PRIOR APPLICATION NUMBER: 60/090694
66	PRIOR FILING DATE: 1998-06-25
67	PRIOR APPLICATION NUMBER: 60/090695
68	PRIOR FILING DATE: 1998-06-25
69	PRIOR APPLICATION NUMBER: 60/090696
70	PRIOR FILING DATE: 1998-06-25
71	PRIOR APPLICATION NUMBER: 60/090862
72	PRIOR FILING DATE: 1998-06-26
73	PRIOR APPLICATION NUMBER: 60/090863





**THIS PAGE BLANK (U3PTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 25, 2002, 10:02:16 : Search time 2767 Seconds  
(without alignments)  
14535.626 Million cell updates/sec

Title: US-09-674-593-1  
Perfect score: 1382  
Sequence: 1 cattatgctaacagcataaa.....taaaagcaattaagtctctgg 1382

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl :

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg\_hum.\*
- 31: em.htg\_inv.\*
- 32: em.htg\_other.\*
- 33: em.htg\_mus.\*
- 34: em.htg\_pln.\*
- 35: em.htg\_rod.\*
- 36: em.htg\_man.\*
- 37: em.htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1382	100.0	1382	9	AF181722	AF181722 Homo sapi
c	1382	100.0	4377	9	AF181720	AF181720 Homo sapi
3	1382	100.0	120029	2	HSJ282H10	AL132672 Homo sapi
4	1380.4	99.9	152966	9	AL359713	AL359713 Human DNA
c	682.4	49.4	1287	9	BC014954	BC014954 Homo sapi
6	595.4	43.1	2167	9	AF181721	AF181721 Homo sapi
c	409.8	29.7	2003	9	AK027036	AK027036 Homo sapi
8	293.4	21.2	6548	9	AB032980	AB032980 Homo sapi
9	252.8	18.3	199957	10	AL589735	AL589735 Mouse DNA
10	86.2	6.2	169103	2	AC125638	AC125638 Rattus no
c	85.4	6.2	139357	9	AC010907	AC010907 Homo sapi
12	75.8	5.5	115756	9	HSJ262L5	AL049795 Human DNA
c	75.8	5.5	322840	2	HSJ2675E8	AL121991 Homo sapi
14	68	4.9	2422	6	AX056382	AX056382 Sequence
c	67.8	4.9	3295	3	AB047038	AB047038 Halocynthia
16	66.2	4.8	141790	10	AL671759	AF429315 Homo sapi
17	61.2	4.4	125020	9	AF429315	AF429315 Homo sapi
c	60.4	4.4	125020	9	AF429315	AF429315 Homo sapi
19	55.4	4.0	77096	2	AC116030	AC116030 Dictyoste
c	55.2	4.0	27057	1	AF169302	AF169302 Burkholde
21	53.6	3.9	12965	1	AE002048	AE002048 Deinococc
c	53.4	3.9	104014	2	AC116921	AC116921 Dictyoste
23	52.8	3.8	47791	2	AC115593	AC115593 Dictyoste
c	52.8	3.8	47791	2	AC116305	AC116305 Dictyoste
25	51.6	3.7	11735	6	AX251838	AX251838 Sequence
c	51.6	3.7	11735	6	AX277870	AX277870 Sequence
27	51.6	3.7	11735	6	AX233553	AX233553 Sequence
c	51.4	3.7	156533	2	AC117070	AC117070 Dictyoste
29	51.2	3.7	182054	10	AL365314	AL365314 Mouse DNA
c	51.2	3.7	184754	2	AC022675	AC022675 Mus muscu
31	50.8	3.7	150613	8	AB001684	AB001684 Chlorella
32	50.8	3.7	175699	2	AC129136	AC129136 Rattus no
c	50.2	3.6	132000	2	AC116976	AC116976 Dictyoste
34	50	3.6	124820	2	AC117073	AC117073 Dictyoste
35	50	3.6	199450	9	AC005674	AC005674 Homo sapi
c	49.4	3.6	20043	2	AC116986	AC116986 Dictyoste
37	49.4	3.6	106144	2	AC116425	AC116425 Dictyoste
c	49.2	3.6	3671	3	AF481922	AF481922 Dictyoste
39	49.2	3.6	33156	2	AC116028	AC116028 Dictyoste
40	49.2	3.6	100998	9	HS343K2	AL031659 Human DNA
41	49.2	3.6	195896	2	AL805976	AL805976 Mus muscu
42	49	3.5	2872	3	DDIUBIRPB	M19492 Slime mold
43	49	3.5	93491	2	AC116967	AC116967 Dictyoste
c	49	3.5	94754	2	AC103188	AC103188 Rattus no
45	49	3.5	173124	2	AC099264	AC099264 Rattus no

ALIGNMENTS

RESULT 1

AF181722

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

AF181722 1382 bp mRNA linear PRI 10-JAN-2000  
Homo sapiens RU2AS (RU2) mRNA, complete cds.

Homo sapiens  
AF181722.1 GI:6684531

Homo sapiens.

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1382)

Van Den Eynde,B.J., Gaugler,B., Probst-Kepper,M., Michaux,L.,

Devuyt,O., Lorge,F., Weynants,P. and Boon,T.

A new antigen recognized by cytolytic T lymphocytes on a human



J. Exp. Med. 190 (12), 1793-1800 (1999)  
20069887  
MEDLINE  
10601354  
PUBMED  
2 (bases 1 to 4377)  
REFERENCE  
Van den Eynde,B.J., Gaugler,B. and Pilotte,L.  
AUTHORS  
Direct Submission  
TITLE  
Submitted (30-AUG-1999) Ludwig Institute for Cancer Research,  
JOURNAL  
Avenue Hippocrate, 74, Brussels 1200, Belgium  
Location/Qualifiers  
FEATURES  
1..4377  
source  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/map="6p22.1"  
/gene="R2"  
1754..>2348  
/gene="R2"  
1754..>2348  
/gene="R2"  
/product="R2"  
complement(<1912..>2166)  
/product="R2AS"  
complement(1912..2166)  
/codon\_start=1  
/product="R2AS"  
/protein\_id="AAF23611.1"  
/db\_xref="GI:6684528"  
/translation="MDDDAARVEGVVAVHKHALHDGLRQVAGPGAAAHLPWPPP  
QLAASREAPLSQRPHTRGAGSPPETNEKLTPQYKER"  
2056..>2348  
/gene="R2"  
/codon\_start=1  
/product="R2"  
/protein\_id="AAF23610.1"  
/db\_xref="GI:6684527"  
/translation="MSGSSARSHLSQPVKSVLYRNGDPFVAGRVRVIEHKVVSF  
EVLKEVIGVQAPFGAVRMIRYPTGRIRKKLDQIQSGNIVAGGGEAPKKL"  
BASE COUNT 1296 a 858 g 828 t 1380 c 15 others  
ORIGIN

Query Match 100.0%; Score 1382; DB 9; Length 4377;  
Best Local Similarity 100.0%; Pred. No. 1.4e-306;  
Matches 1382; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy - 1 CATTATGCTAACAGCATAAACATGCGGGGTGGAGCAGGGTGCACAAAAGTGAAGTGTG 60  
Db 2903 CATTATGCTAACAGCATAAACATGCGGGGTGGAGCAGGGTGCACAAAAGTGAAGTGTG 2844

Qy 61 TCAATTTCTACTTGGAAAGAGTTGAAATTAATTAACAGTACGGGAAATGCAGAGCAA 120  
Db 2843 TCAATTTCTACTTGGAAAGAGTTGAAATTAATTAACAGTACGGGAAATGCAGAGCAA 2784

Qy 121 TTTTCTCTCTGGTGACAAATATAGTGTCAACACTTGGAACTGATTTTAAAGAAATGTTTA 180  
Db 2783 TTTTCTCTCTGGTGACAAATATAGTGTCAACACTTGGAACTGATTTTAAAGAAATGTTTA 2724

Qy 181 TTTAAATTTAAAGGATGGATTCTCAAGGAAAGAAATTAAGGAAAGGAAAGAAAGAAAGT 240  
Db 2723 TTTAAATTTAAAGGATGGATTCTCAAGGAAAGAAATTAAGGAAAGGAAAGAAAGAAAGT 2664

Qy 241 AACAGAAACCCAAAGTATCAGTTTGTGTCACACTAACCTTTGCAAGGATACCTTTTATTT 300  
Db 2663 AACAGAAACCCAAAGTATCAGTTTGTGTCACACTAACCTTTGCAAGGATACCTTTTATTT 2604

Qy 301 TCTTTAAGATCTCTGTTGTTTATACAGAGTTTAAAGTTTACTCTACTGCTGACCCCAAG 360  
Db 2603 TCTTTAAGATCTCTGTTGTTTATACAGAGTTTAAAGTTTACTCTACTGCTGACCCCAAG 2544

Qy 361 TGAATTCCTCTCTCAGTCAAGTGTCAACCTCTACCCCACTGCAACAGAGAGTTTGTG 420  
Db 2543 TGAATTCCTCTCTCAGTCAAGTGTCAACCTCTACCCCACTGCAACAGAGAGTTTGTG 2484

Qy 421 AGGGGCATCAATCAACCCGAGAGTACAGCCCTCAACCACTCAGGTGTGGGGGGTAG 480  
|||||

Db 2483 AGGGGCATCAATCAACCCGAGAGTACAGCCCTCAACCACTCAAGTGTGGGGGGTAG 2424

Qy 481 GGATCTGCATTCTTCATATCAACCCCACTATAGGCACCTAAATGGGTGGGGGGTGG 540  
|||||

Db 2423 GGATCTGCATTCTTCATATCAACCCCACTATAGGCACCTAAATGGGTGGGGGGTGG 2364  
|||||

Qy 541 GGGAGACCGACTCACTTTCAGTTTCTTGAAGCTTCTTGGCTTCCACACCTAAATTTGCC 600  
|||||

Db 2363 GGGAGACCGACTCACTTTCAGTTTCTTGAAGCTTCTTGGCTTCCACACCTAAATTTGCC 2304  
|||||

Qy 601 CCGCTCTGGATCTGGTCTAGCTTCGGGATTCGGTGGCCAGTCCCGGGGTGTAGATGTC 660  
|||||

Db 2303 CCGCTCTGGATCTGGTCTAGCTTCGGGATTCGGTGGCCAGTCCCGGGGTGTAGATGTC 2244  
|||||

Qy 661 CTGAGGCCCAAGGGGTGCCTGAACGCCGGGTCACTCTCTTCAGGAAGACTTTCGAAG 720  
|||||

Db 2243 CTGAGGCCCAAGGGGTGCCTGAACGCCGGGTCACTCTCTTCAGGAAGACTTTCGAAG 2184  
|||||

Qy 721 CTGAGACCTTCTTCTCATGTGATGACGACGGCGGCCCGGTGTAGAGGGTTCGCCGTG 780  
|||||

Db 2183 CTGAGACCTTCTTCTCATGTGATGACGACGGCGGCCCGGTGTAGAGGGTTCGCCGTG 2124  
|||||

Qy 781 CCGTACACAAGACGCTCTTCACGACGGGTGAGACAGGTGGCTGGACTGGCGTGTG 840  
|||||

Db 2123 CCGTACACAAGACGCTCTTCACGACGGGTGAGACAGGTGGCTGGACTGGCGTGTG 2064  
|||||

Qy 841 CCGCTCATCTTCCCGCTGGCGGCCCTCAGCTCGCTTTCGGCTCGGAGGGCACCTC 900  
|||||

Db 2063 CCGCTCATCTTCCCGCTGGCGGCCCTCAGCTCGCTTTCGGCTCGGAGGGCACCTC 2004  
|||||

Qy 901 CGCTGTCCAGCGGCTCACCGACCCAGCGGGGATCGCTCTCTGAAAGAAACGAGA 960  
|||||

Db 2003 CGCTGTCCAGCGGCTCACCGACCCAGCGGGGATCGCTCTCTGAAAGAAACGAGA 1944  
|||||

Qy 961 AACTGACCAATCCACAGGTGAAAGAGAGTAAAGCCCTGCGCTAGCGTCCACCCAGA 1020  
|||||

Db 1943 AACTGACCAATCCACAGGTGAAAGAGAGTAAAGCCCTGCGCTAGCGTCCACCCAGA 1884  
|||||

Qy 1021 GGAGACACTAGGAGCTTCAGGACTCGCAGTAGAGCTCAAGTTTTCACCGTGGCGTGC 1080  
|||||

Db 1883 GGAGACACTAGGAGCTTCAGGACTCGCAGTAGAGCTCAAGTTTTCACCGTGGCGTGC 1824  
|||||

Qy 1081 ACAGCCAATCAGGACCCGCTGCGGCACCAACAGGTTCACCTGTACGGGCGAGAT 1140  
|||||

Db 1823 ACAGCCAATCAGGACCCGCTGCGGCACCAACAGGTTCACCTGTACGGGCGAGAT 1764  
|||||

Qy 1141 CAAGGTGACAGCTTCTCAGCAGGAGCCGGAACCGCGGGGCTTCAACAGGCGACGCC 1200  
|||||

Db 1763 CAAGGTGACAGCTTCTCAGCAGGAGCCGGAACCGCGGGGCTTCAACAGGCGACGCC 1704  
|||||

Qy 1201 TAGTGAGGCGAGGAGAGGAGGAGGACACACACACACACACAAATATGTTGAAACC 1260  
|||||

Db 1703 TAGTGAGGCGAGGAGAGGAGGAGGAGGACACACACACACACAAATATGTTGAAACC 1644  
|||||

Qy 1261 CAATTTCTTACATCATATCTGTGTACCTTTTCCAAACAGCCTAAATTTTCTTTCTCTC 1320  
|||||

Db 1643 CAATTTCTTACATCATATCTGTGTACCTTTTCCAAACAGCCTAAATTTTCTTTCTCTC 1584  
|||||

Qy 1321 TCTTTCACCTTTACCCCTCAATCTCTGTCTTCTCCCAAAATTAAGCAATTAAGTCTCT 1380  
|||||

Db 1583 TCTTTCACCTTTACCCCTCAATCTCTGTCTTCTCCCAAAATTAAGCAATTAAGTCTCT 1524  
|||||

Qy 1381 GG 1382  
|||

Db 1523 GG 1522

RESULT 3  
HSJ282H10  
LOCUS  
DEFINITION  
Homo sapiens chromosome 6 clone RP1-282H10 map p22.1-22.3, \*\*\*  
SEQUENCING IN PROGRESS \*\*\*, 5 unordered pieces.  
ACCESSION  
AL132672



Qy 961 AACTGACGAATCCACAGGTGAAGAGAGAGTAACGGCCGCTGCGCTAGCGGTCCACCCAG 1020  
 Db 22008 AACTGACGAATCCACAGGTGAAGAGAGAGTAACGGCCGCTGCGCTAGCGGTCCACCCAG 22067  
 Qy 1021 GGAGACACTAGGAGCTTCAGGACTCGGAGTAGAGGCTCAAGTTTTCACCGTGGCGTGC 1080  
 Db 22068 GGAGACACTAGGAGCTTCAGGACTCGGAGTAGAGGCTCAAGTTTTCACCGTGGCGTGC 22127  
 Qy 1081 ACAGCCAATCAGGACCCGCGAGTGGCGGACACACACAGGTTTCACCTGCTACGGGCGAGAT 1140  
 Db 22128 ACAGCCAATCAGGACCCGCGAGTGGCGGACACACACAGGTTTCACCTGCTACGGGCGAGAT 22187  
 Qy 1141 CAAGGTGACAGCTTCGTGAGCAGAGCGCGGAACGCGGGGCGCTTCAACAGGACGCC 1200  
 Db 22188 CAAGGTGACAGCTTCGTGAGCAGAGCGCGGAACGCGGGGCGCTTCAACAGGACGCC 22247  
 Qy 1201 TAGTGAGGCGAGGAGGAGGAGGACGACACACACACACACACAATATGGTCAACCC 1260  
 Db 22248 TAGTGAGGCGAGGAGGAGGAGGACGACACACACACACACAATATGGTCAACCC 22307  
 Qy 1261 CAATTTCTTACATCATCTCTGCTACCTTTCCAAACAGGCTTAATTTTCTTTCTCTC 1320  
 Db 22308 CAATTTCTTACATCATCTCTGCTACCTTTCCAAACAGGCTTAATTTTCTTTCTCTC 22367  
 Qy 1321 TTCTTGACCTTTACCCCTCAATCTCTGCTCTCTCCCAAAATTAAGCAATTAAGTTCC 1380  
 Db 22368 TTCTTGACCTTTACCCCTCAATCTCTGCTCTCTCCCAAAATTAAGCAATTAAGTTCC 22427  
 Qy 1381 GG 1382  
 Db 22428 GG 22429

RESULT 4  
 AL359713  
 LOCUS Human DNA sequence from clone RP11-95P3 on chromosome 6, complete  
 DEFINITION  
 ACCESSION AL359713  
 VERSION AL359713.25 GI:13938809  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 152966)  
 Kimberley, A.  
 Direct Submission  
 Submitted (01-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
 requests: clonerequest@sanger.ac.uk  
 On May 3, 2001 this sequence version replaced gi:13446455.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest. The following  
 abbreviations are used to associate primary accession numbers given  
 in the feature table with their source databases: Em:, EMBL; Sw:,  
 SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP  
 database can be found at  
 http://www.sanger.ac.uk/projects/C\_elegans/wormpep This sequence  
 was generated from part of bacterial clone contigs of human  
 chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping  
 Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr6  
 RP11-95P3 is from the library RPI-11.1 constructed by the group of  
 Pieter de Jong. For further details see  
 http://www.chori.org/bacpac/home.htm  
 VECTOR: pBACe3.6

IMPORTANT: This sequence is not the entire insert of clone  
 RP11-95P3 it may be shorter because we sequence overlapping  
 sections only once, except for a 100 base overlap.  
 The true left end of clone RP11-95P3 is at 1 in this sequence. The  
 true left end of clone RP1-73M23 is at 152K67 in this sequence. The  
 true right end of clone RP11-40E20 is at 17700 in this sequence.

## FEATURES

Location/Qualifiers  
 1..152966  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /clone="RP11-95P3"  
 /clone\_lib="RPI-11.1"  
 19..544  
 /note="match: GSS: Em:AZ517849"  
 24..379  
 /note="match: GSS: Em:B63526"  
 32..494  
 /note="match: GSS: Em:AQ315706"  
 70..589  
 /note="match: GSS: Em:AQ285677"  
 179..362  
 /note="MIR repeat: matches 65..262 of consensus"  
 521..582  
 /note="L2 repeat: matches 2683..2741 of consensus"  
 720..1022  
 /note="Alusx repeat: matches 1..303 of consensus"  
 1327..1448  
 /note="AluJo/FRAM repeat: matches 181..298 of consensus"  
 2264..2575  
 /note="Alusq repeat: matches 1..311 of consensus"  
 3014..3296  
 /note="match: STS: Em:G27290"  
 3014..3093  
 /note="40 copies 2 mer ga 75% conserved"  
 5024..5477  
 /note="L1M1 repeat: matches 1012..1598 of consensus"  
 5581..5859  
 /note="Alusg repeat: matches 18..294 of consensus"  
 5986..6359  
 /note="L1MEC repeat: matches 2228..2259 of consensus"  
 6367..6515  
 /note="FLAM\_C repeat: matches 1..143 of consensus"  
 6971..7198  
 /note="L2 repeat: matches 1512..1735 of consensus"  
 8403..8838  
 /note="LTR7 repeat: matches 1..445 of consensus"  
 8799..9441  
 /note="match: GSS: Em:B54691"  
 complement(9393..9853)  
 /note="match: GSS: Em:AQ285817"  
 10311..10437  
 /note="L2 repeat: matches 2612..2747 of consensus"  
 10710..11022  
 /note="AluJo repeat: matches 1..309 of consensus"  
 12006..12156  
 /note="L2 repeat: matches 2594..2748 of consensus"  
 12182..13011  
 /note="L2 repeat: matches 1374..2344 of consensus"  
 13404..13631  
 /note="L14 copies 2 mer at 55% conserved"  
 13479..13630  
 /note="38 copies 4 mer tata 60% conserved"  
 13785..14820  
 /note="L1ME repeat: matches 4766..5798 of consensus"  
 14928..16008  
 /note="L1PB3 repeat: matches 5106..6149 of consensus"  
 16029..16110

```
repeat_region /note="L1PB3 repeat: matches 5054. .5133 of consensus"
16428. .16829 /note="L1MC3 repeat: matches 6189. .6603 of consensus"
repeat_region /note="L1MC3 repeat: matches 6189. .6603 of consensus"
16868. .17150 /note="AluX repeat: matches 1. .285 of consensus"
repeat_region /note="AluX repeat: matches 1. .285 of consensus"
17246. .17991 /note="L1MC4 repeat: matches 6608. .7343 of consensus"
repeat_region /note="L1MC4 repeat: matches 6608. .7343 of consensus"
17992. .18675 /note="MER67C repeat: matches 1. .710 of consensus"
repeat_region /note="MER67C repeat: matches 1. .710 of consensus"
18676. .18729 /note="L1MC4 repeat: matches 7343. .7396 of consensus"
repeat_region /note="L1MC4 repeat: matches 7343. .7396 of consensus"
18731. .19458 /note="MER4D repeat: matches 232. .973 of consensus"
misc_feature complement(1929. .19542)
/note="match: GSS: Em:AQ342961"
misc_feature complement(19031. .19542)
/note="match: GSS: Em:AQ053552"
misc_feature complement(19065. .19495)
/note="match: GSS: Em:AQ545892"
misc_feature complement(19130. .19546)
/note="match: GSS: Em:B94529"
19265. .19495 /note="match: GSS: Em:AQ373351 Em:AQ375803
/note="match: GSS: Em:G59266"
misc_feature complement(19269. .19546)
/note="match: GSS: Em:AQ663911"
misc_feature complement(19325. .19546)
/note="match: GSS: Em:AQ021494"
misc_feature complement(19337. .19501)
/note="match: GSS: Em:AQ268095"
misc_feature complement(19370. .19546)
/note="match: GSS: Em:AQ128461"
19566. .19739 /note="AluSg/x repeat: matches 126. .295 of consensus"
repeat_region /note="AluSg/x repeat: matches 126. .295 of consensus"
19740. .20100 /note="L1MD3 repeat: matches 7391. .7739 of consensus"
repeat_region /note="L1MD3 repeat: matches 7391. .7739 of consensus"
21121. .21429 /note="AluJo repeat: matches 1. .310 of consensus"
repeat_region /note="AluJo repeat: matches 1. .310 of consensus"
22589. .22728 /note="70 copies 2 mer aa 60% conserved"
repeat_region /note="70 copies 2 mer aa 60% conserved"
22591. .22670 /note="20 copies 4 mer aaag 80% conserved"
repeat_region /note="20 copies 4 mer aaag 80% conserved"
22686. .22761 /note="19 copies 4 mer aagg 93% conserved"
repeat_region /note="19 copies 4 mer aagg 93% conserved"
23098. .23393 /note="AluNb repeat: matches 4. .298 of consensus"
repeat_region /note="AluNb repeat: matches 4. .298 of consensus"
23468. .23578 /note="L2 repeat: matches 2580. .2700 of consensus"
repeat_region /note="L2 repeat: matches 2580. .2700 of consensus"
23980. .24292 /note="AluSg repeat: matches 2. .310 of consensus"
misc_feature complement(2708. .27677)
/note="match: GSS: Em:AQ702871"
repeat_region /note="AluNb repeat: matches 163. .299 of consensus"
27819. .27957 /note="MLTIC repeat: matches 32. .301 of consensus"
repeat_region /note="MLTIC repeat: matches 32. .301 of consensus"
29080. .29171 /note="MLTIIH repeat: matches 433. .526 of consensus"
repeat_region /note="MLTIIH repeat: matches 433. .526 of consensus"
29303. .29711 /note="L2 repeat: matches 2323. .2710 of consensus"
repeat_region /note="L2 repeat: matches 2323. .2710 of consensus"
29712. .29834 /note="MLTIG repeat: matches 29. .147 of consensus"
repeat_region /note="MLTIG repeat: matches 29. .147 of consensus"
30140. .30547 /note="L2 repeat: matches 1806. .2252 of consensus"
misc_feature complement(30720. .31207)
/note="match: STS: Em:HS1J17T"
repeat_region /note="MER58A repeat: matches 1. .208 of consensus"
30919. .31120 /note="PHEIC repeat: matches 1. .371 of consensus"
repeat_region /note="PHEIC repeat: matches 1. .371 of consensus"
32109. .32472 /note="MIR repeat: matches 98. .250 of consensus"
repeat_region /note="MIR repeat: matches 98. .250 of consensus"
33360. .33511 /note="MIR repeat: matches 98. .250 of consensus"
repeat_region /note="MIR repeat: matches 98. .250 of consensus"
34031. .34331
```

```
repeat_region /note="AluX repeat: matches 1. .297 of consensus"
34621. .34732 /note="MIR repeat: matches 120. .232 of consensus"
repeat_region /note="MIR repeat: matches 120. .232 of consensus"
34888. .35007 /note="L2 repeat: matches 1836. .1959 of consensus"
repeat_region /note="L2 repeat: matches 1836. .1959 of consensus"
35164. .35461 /note="AluSg repeat: matches 2. .298 of consensus"
repeat_region /note="AluSg repeat: matches 2. .298 of consensus"
35532. .35990 /note="L2 repeat: matches 2254. .2710 of consensus"
repeat_region /note="L2 repeat: matches 2254. .2710 of consensus"
36631. .36790 /note="L2 repeat: matches 2342. .2501 of consensus"
repeat_region /note="L2 repeat: matches 2342. .2501 of consensus"
37251. .37544 /note="AluX repeat: matches 1. .294 of consensus"
repeat_region /note="AluX repeat: matches 1. .294 of consensus"
38446. .38493 /note="12 copies 4 mer caca 75% conserved"
repeat_region /note="12 copies 4 mer caca 75% conserved"
38467. .38492

Query Match 99.9%; Score 1380.4; DB 9; Length 152966;
Best Local Similarity 99.9%; Pred. No. 6.8e-306;
Matches 1381; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CATTATGCTAACACATAACATGCAGGGGTGGGAGCAGGTCACAAAAGTGAGTGTG 60
Db 53422 CATTATGCTAACACATAACATGCAGGGGTGGGAGCAGGTCACAAAAGTGAGTGTG 53481
QY 61 TCAATTCCTACTTGAATGAAAGTTGAAATTAATTTAAACAGTACGGGAAATGCAGAGCAA 120
Db 53482 TCAATTCCTACTTGAATGAAAGTTGAAATTAATTTAAACAGTACGGGAAATGCAGAGCAA 53541
QY 121 TTTTCTCTCTGGTGACAATATAGTGTCCAACACTTGAAGTGATTTTAAAGATGTTTA 180
Db 53542 TTTTCTCTCTGGTGACAATATAGTGTCCAACACTTGAAGTGATTTTAAAGATGTTTA 53601
QY 181 TTTAAATTAAGAGTGGATTTCCAGGAAAGAAATAAGGAAAGAGAAAGAAAGAACTG 240
Db 53602 TTTAAATTAAGAGTGGATTTCCAGGAAAGAAATAAGGAAAGAGAAAGAAAGAACTG 53661
QY 241 AACGAAAGCGAAAGATATCAGTTTGTCTACTTAACCTTTGCAAGGATACCTTTTATTT 300
Db 53662 AACGAAAGCGAAAGATATCAGTTTGTCTACTTAACCTTTGCAAGGATACCTTTTATTT 53721
QY 301 TCTTTAAGATTCCTGTTTATACACAGATTTTAAAGTTTACTCTCTACTGTCACCAAG 360
Db 53722 TCTTTAAGATTCCTGTTTATACACAGATTTTAAAGTTTACTCTCTACTGTCACCAAG 53781
QY 361 TGAATTCCTTCTCCAGTCACAGTGTCAACCTCTACCCCTCACTGCAACGAGATTTTG 420
Db 53782 TGAATTCCTTCTCCAGTCACAGTGTCAACCTCTACCCCTCACTGCAACGAGATTTTG 53841
QY 421 AGGGGCATCAATCACACCGAGAAATCAGCCCTCAACCACTGAGGTGTGGGGGGGTAG 480
Db 53842 AGGGGCATCAATCACACCGAGAAATCAGCCCTCAACCACTGAGGTGTGGGGGGGTAG 53901
QY 481 GGATCTGATTTCTTCATATCAACCCACACTATAGGACACCTAAATGGTGGGGGTGG 540
Db 53902 GGATCTGATTTCTTCATATCAACCCACACTATAGGACACCTAAATGGTGGGGGTGG 53961
QY 541 GGGAGACGACTCACTTGAGTTTCTTGAAGGCTTCTGGCTTCCAGCCACGTAATTCGCC 600
Db 53962 GGGAGACGACTCACTTGAGTTTCTTGAAGGCTTCTGGCTTCCAGCCACGTAATTCGCC 54021
QY 601 CGCTCTGGATCTGTTAGCTTTCGGATTCGGTGCACAGTCCCGGGGTGTAGATGTTTC 660
Db 54022 CGCTCTGGATCTGTTAGCTTTCGGATTCGGTGCACAGTCCCGGGGTGTAGATGTTTC 54081
QY 661 CTGAGCGCCCCAAAGGGTGCCTGACGCGCGCGGTACCTCTCTCAGGAAGACTTCGAAG 720
Db 54082 CTGAGCGCCCCAAAGGGTGCCTGACGCGCGCGGTACCTCTCTCAGGAAGACTTCGAAG 54141
QY 721 CTGGACACCTCTCTCTCATGATGACGAGCGCGCGCGCGGTAGAAAGGGTCCCGGTTG 780
Db 54142 CTGGACACCTCTCTCTCATGATGACGAGCGCGCGCGCGGTAGAAAGGGTCCCGGTTG 54201
```

781 CGGTACAAAGCAGCGCTTTCACGACGGGCTGAGACAGGTGGTGGACCTGGCGCTGCTG 840  
Db 54202 CGGTACAAAGCAGCGCTTTCACGACGGGCTGAGACAGGTGGTGGACCTGGCGCTGCTG 54261  
Qy 841 CGGCTCATCTTCCCGCTGGCGCGGCTCAGCTCGCTGCTTGGCTGCGGAGGACCTC 900  
Db 54262 CGGCTCATCTTCCCGCTGGCGCGGCTCAGCTCGCTGCTTGGCTGCGGAGGACCTC 54321  
Qy 901 CGGTGTCACAGCGGCTTTCACGACCGGCTGAGACAGGTGGTGGACCTGGCGCTGCTG 960  
Db 54322 CGGTGTCACAGCGGCTTTCACGACCGGCTGAGACAGGTGGTGGACCTGGCGCTGCTG 54381  
Qy 961 AACTGACGAATCCACAGGTGAAGAGAGTAACGCGCGCTGAGCGCTGACCGCAGCA 1020  
Db 54382 AACTGACGAATCCACAGGTGAAGAGAGTAACGCGCGCTGAGCGCTGACCGCAGCA 54441  
Qy 1021 GGAGACACTAGGAGCTTTCAGAGGCTGGAGTAGAGCTCAAGTTTTCACCGTGGCGTGC 1080  
Db 54442 GGAGACACTAGGAGCTTTCAGAGGCTGGAGTAGAGCTCAAGTTTTCACCGTGGCGTGC 54501  
Qy 1081 ACAGCCAAATCAGGACCGGCTGCGGCGCAGCCACACAGGTTCACCTGCTACGGGCA 1140  
Db 54502 ACAGCCAAATCAGGACCGGCTGCGGCGCAGCCACACAGGTTCACCTGCTACGGGCA 54561  
Qy 1141 CAAGGTGGACAGCTTTCAGAGGAGCGGAGCGGCGGCTTCAACAGAGCAGCGC 1200  
Db 54562 CAAGGTGGACAGCTTTCAGAGGAGCGGAGCGGCGGCTTCAACAGAGCAGCGC 54621  
Qy 1201 TAGTGAGGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1260  
Db 54622 TAGTGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 54681  
Qy 1261 CAATTTCTTACATCATATCTGCTGCTACCTTTCACACAGCCTAAATTTTCTCTC 1320  
Db 54682 CAATTTCTTACATCATATCTGCTGCTACCTTTCACACAGCCTAAATTTTCTCTC 54741  
Qy 1321 TTCTGACCTTTACCCCTCAATCTCTGCTCTCTCTCCCAATTAAGTTTCCT 1380  
Db 54742 TTCTGACCTTTACCCCTCAATCTCTGCTCTCTCTCCCAATTAAGTTTCCT 54801  
Qy 1381 GG 1382  
Db 54802 GG 54803

RESULT 5  
BC014954/c  
LOCUS BC014954  
DEFINITION Homo sapiens, clone MGC:22980 IMAGE:4874845, mRNA, complete cds.  
ACCESSION BC014954  
VERSION BC014954.1 GI:15928979  
KEYWORDS MGC.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1287)  
Strausberg, R.  
Direct Submission  
Submitted (01-OCT-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Genome Sequence Centre,  
BC Cancer Agency, Vancouver, BC, Canada  
info@bcgsc.bc.ca  
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,

Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,  
Letitia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo  
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven  
Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline  
Schein, Duane Smalls, Michael Smith, Lorraine Spence, Jeff Scott,  
Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,  
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAL Plate: 34 Row: m Column: 13  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 7706690.

FEATURES  
Location/Qualifiers  
1..1287  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="MGC:22980 IMAGE:4874845"  
/tissue\_type="Eye, normal, pigmented retinal epithelium"  
/clone\_lib="NIH\_MGC\_43"  
/lab\_host="DH10B-R"  
/note="Vector: pOTB7"  
181..984  
CDS  
/codon\_start=1  
/product="Unknown (protein for MGC:22980)"  
/protein\_id="AAH14954.1"  
/db\_xref="GI:15928980"  
/translation="MGAVSAGEDNDLDLNLEAQTIRVEQAFRAIDRDGYILE  
GYFDNAYKDLAKHGHILSAPIYSEVMEALKLPGLSLFNLGSGTGLSTWVGIL  
GPGFNIHGIHSDLVVEYAEKLESFKNSDFKRSALGAVRLPGRCLPTRSS  
ELRRPAGKSGSSSHLSQPVKSLVYRNGDPFYAGRRVVIHEKVVSPFEVLK  
EVTGGVQAPFGAVRNITPTGHRIRKLDQIQSGGNYVAGGQEAFFKLK"

BASE COUNT 343 a 281 c 359 g 304 t  
ORIGIN

Query Match 49.4%; Score 682.4; DB 9; Length 1287;  
Best Local Similarity 99.9%; Pred. No. 5.3e-146;  
Matches 683; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 264 TTTGGTCACATAACCTTTGCAAGGATACCTTTTATTTTCTTAAAGATCTCTGTTGTTAT 323  
Db 1272 TTTGGTCACATAACCTTTGCAAGGATACCTTTTATTTTCTTAAAGATCTCTGTTGTTAT 1213  
Qy 324 ACACAGATTTTAAAGTTTACTCTACTGTCACCAAGTAAATTCCTTCCAGTACAG 383  
Db 1212 ACACAGATTTTAAAGTTTACTCTACTGTCACCAAGTAAATTCCTTCCAGTACAG 1153  
Qy 384 TGTCAACCTCTACCCCAACTGCAAGAGAGTTTGGAGGGGCAATCAATCACACCGAGAA 443  
Db 1152 TGTCAACCTCTACCCCAACTGCAAGAGAGTTTGGAGGGGCAATCAATCACACCGAGAA 1093  
Qy 444 GTCACAGCCCTCAACCACTGAGGTGTGGGGGGTAGGGATCTCCATTTCTTATATCAA 503  
Db 1092 GTCACAGCCCTCAACCACTGAGGTGTGGGGGGTAGGGATCTCCATTTCTTATATCAA 1033  
Qy 504 CCCACACTATAGGACACCTAAATGGTGGCGGTGGGGAGACCGACTCACCTTGAGTTT 563  
Db 1032 CCCACACTATAGGACACCTAAATGGTGGCGGTGGGGAGACCGACTCACCTTGAGTTT 973  
Qy 564 CTTGAAGGCTTCTGGCCTCCAGCCAGTAAATGGCCCCGCTCTGGATCTGGTCTAGCTT 623  
Db 972 CTTGAAGGCTTCTGGCCTCCAGCCAGTAAATGGCCCCGCTCTGGATCTGGTCTAGCTT 913  
Qy 624 CCGGATTCGGTGGCGAGTCCGGGGGTAGATGTTCTTACCGCCCAAGGGTGGCTG 683  
Db 912 CCGGATTCGGTGGCGAGTCCGGGGGTAGATGTTCTTACCGCCCAAGGGTGGCTG 853  
Qy 684 AACCGCGCGGTTCACCTTCTTACAGAGACTTCGAAGCTGACACACCTTCTTCATGAT 743  
Db 852 AACCGCGCGGTTCACCTTCTTACAGAGACTTCGAAGCTGACACACCTTCTTCATGAT 793  
Qy 744 GAGCAGCGCGCGCGCGGTAGAGGGGTCCCGGTGCGGTACACAGCAGCTTCTTCAC 803

Db	792	GACGACGCGCGCCCGGTAGAGAGGGTCCCGCTTCGGGTACACAAAGCAGCTCTTCAC	733
Qy	804	GACGGCTGACAGAGTGGTGGACCTGGCGCTGCTCCGCTATCTTCCTCCGCTGGCCG	863
Db	732	GACGGCTGACAGAGTGGTGGACCTGGCGCTGCTCCGCTATCTTCCTCCGCTGGCCG	673
Qy	864	CCGCTCAGCTCGCTGCTCGCGTGGGAGGCACTCCGCTGTGCCAGCGCTTCACCGC	923
Db	672	CCGCTCAGCTCGCTGCTCGCGTGGGAGGCACTCCGCTGTGCCAGCGCTTCACCGC	613
Qy	924	ACCCAGGCGCGGATCGCTCTCT	947
Db	612	ACCCAGGCGCGGATCGCTCTCT	589
RESULT 6			
LOCUS	AF181721/c	2167 bp	mrna linear PRI 10-JAN-2000
DEFINITION	Homo sapiens RU2S (RU2) mRNA, complete cds.		
ACCESSION	AF181721		
VERSION	AF181721.1	GI:6684529	
KEYWORDS	Homo sapiens.		
SOURCE	Homo sapiens		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	Van Den Eynde, B.J., Gaugler, B., Probst-Keppler, M., Michaux, L., Devuyt, O., Lorge, F., Weynants, P., and Boon, T.		
TITLE	A new antigen recognized by cytolytic T lymphocytes on a human kidney tumor results from reverse strand transcription		
JOURNAL	J. Exp. Med. 190 (12), 1793-1800 (1999)		
MEDLINE	20069887		
PUBMED	10601354		
REFERENCE	2 (bases 1 to 2167)		
AUTHORS	Van den Eynde, B.J., Gaugler, B. and Pilotte, L.		
TITLE	Direct Submission		
JOURNAL	Submitted (30-AUG-1999) Ludwig Institute for Cancer Research, Avenue Hippocrate, 74, Brussels 1200, Belgium		
FEATURES	Location/Qualifiers		
source	1..2167		
gene	/organism="Homo sapiens"		
CDS	/db_xref="taxon:9606"		
	/chromosome="6"		
	/map="6p22.1"		
	1..2167		
	/gene="RU2"		
	303..1733		
	/gene="RU2"		
	/codon_start=1		
	/product="RU2S"		
	/protein_id="AAP23612.1"		
	/db_xref="GI:6684530"		
	/translation="MSGSSARSHLSQPVVYVLRNGDPFFVAGRRVVIHERKVSFS		
	EVLKEVTGVQPFAGVRIYPTGHRIRKLDQISGNGYVAGGAEAFKLNLYDI		
	GEIKRPMVEYVNVKVIHRSINVSFRKPLQEPCTIFLIANGDLINPASILIPR		
	KTLNWDHVLQWYTEKTLRSGAVHRLYTLGLKVESGALENGQFVYVGRDKFKL		
	PYGELLDKSTMRPFGKASLLPIVGRSKSGNDRHSKSTVGSSDNSPQPLKR		
	KGRKEDVSRLLTKQNLKNSQETIPNSDEGIFKAGARSETRGAARVQDEDTQ		
	VEVPDQRAIVDEEEDGKANKAEQDIPNSDEGIFKAGARSETRGAARVQDEDTQ		
	DHSDQARPAIVNGTDENGEELQVNNELQLVLDKERKSGAGSGQDEADVDPPRP		
	PRPEVKITSPENNNQNDYAAVA		
BASE COUNT	675 a	440 c	576 g 476 t
ORIGIN			
Query Match	43.1%	Score 595.4;	DB 9; Length 2167;
Best Local Similarity	99.8%	Pred. No. 5.7e-126;	
Matches	596;	Conservative 0;	Mismatches 1; Indels 0; Gaps 0;
Qy	554	ACTTGAGTTCTTGAGGCTTCCTGGCGCTCCAGCCACGTAATTCGCCCGCTCTGGATCT	613
Db	597	AATTGAGTTCTTGAGGCTTCCTGGCGCTCCAGCCACGTAATTCGCCCGCTCTGGATCT	538

Qy	614	GGTCTAGCTTCCGGATTTCGGTGGCCAGTCCGCGGGTGTAGATGTTCTCTCAGCGCCCAA	673
Db	537	GGTCTAGCTTCCGGATTTCGGTGGCCAGTCCGCGGGTGTAGATGTTCTCTCAGCGCCCAA	478
Qy	674	AGGTGCTCAAGCGCCGCTCAGCTTCCTTTCAGGAAGACTTCGAAGCTGACACCTTCT	733
Db	477	AGGTGCTCAAGCGCCGCTCAGCTTCCTTTCAGGAAGACTTCGAAGCTGACACCTTCT	418
Qy	734	TCTCATGGATGACGACGCGCGCCCGCTAGAGAGGGTCCCGCTTCGGTTCACAAAGCA	793
Db	417	TCTCATGGATGACGACGCGCGCCCGCTAGAGAGGGTCCCGCTTCGGTTCACAAAGCA	358
Qy	794	CGCTTTTCACGAGCGGCTGAGACAGGTGGTGGACCTGGCGCTGCTCGCGTCTATCTTC	853
Db	357	CGCTTTTCACGAGCGGCTGAGACAGGTGGTGGACCTGGCGCTGCTCGCGTCTATCTTC	298
Qy	854	CGCTGGCGCGCTCAGCTCGCTGCTCGGTCGCGGAGGACCTCCGCTGTCCAGCG	913
Db	297	CGCTGGCGCGCTCAGCTCGCTGCTCGGTCGCGGAGGACCTCCGCTGTCCAGCG	238
Qy	914	GCCTCACCGCACCCAGCGCGGATCGCTTCTTGAACGAACGAGAACTGACGAATCC	973
Db	237	GCCTCACCGCACCCAGCGCGGATCGCTTCTTGAACGAACGAGAACTGACGAATCC	178
Qy	974	ACAGGTGAAAGAGAAAGTAACGCGCGCTAGGCTGCACCCAGAGGAGACACTAGGA	1033
Db	177	ACAGGTGAAAGAGAAAGTAACGCGCGCTAGGCTGCACCCAGAGGAGACACTAGGA	118
Qy	1034	GCCTGAGGAGTCCGAGTACAGCTCAAGTTTTCACCGTGGCGTGCACGCCAATCAGG	1093
Db	117	GCCTGAGGAGTCCGAGTACAGCTCAAGTTTTCACCGTGGCGTGCACGCCAATCAGG	58
Qy	1094	ACCGCGAGTGGCGCACACACAGGTTCACCTGCTACGGGCGAGANTCAAGGTGGAC	1150
Db	57	ACCGCGAGTGGCGCACACACAGGTTCACCTGCTACGGGCGAGANTCAAGGTGGAC	1
RESULT 7			
LOCUS	AK027036/c	2003 bp	mrna linear PRI 29-SEP-2000
DEFINITION	Homo sapiens cDNA: FLJ23383 fis, clone HEP16466, highly similar to AF181721 Homo sapiens RU2S mRNA.		
ACCESSION	AK027036		
VERSION	AK027036.1	GI:10440050	
KEYWORDS	oligo capping; fis (full insert sequence).		
SOURCE	Homo sapiens hepatoma cell_line:HepG2 cDNA to mRNA, clone_lib:HEP clone:HEP16466.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	Kawabata, A., Hikiji, T., Kobatake, N., Inagaki, H., Ikema, Y., Okamoto, S., Ohtani, R., Ota, T., Suzuki, Y., Okayashi, M., Nishi, T., Shibahara, T., Tanaka, T., Nakamura, Y., Isogai, T. and Sugano, S.		
TITLE	NEDO human cDNA sequencing project		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 2003)		
AUTHORS	Sugano, S., Suzuki, Y., Ota, T., Okayashi, M., Nishi, T., Isogai, T., Shibahara, T., Tanaka, T. and Nakamura, Y.		
TITLE	Direct Submission		
JOURNAL	Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna@leims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)		
COMMENT	NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing; Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).		
FEATURES	Location/Qualifiers		



```
source 1. .2003
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEP18466"
/cell_line="HepG2"
/tissue_type="hepatoma"
/clone_lib="HEP"
/notes="cloning vector pME18SFL3"
misc_feature 1. .2003
/notes="highly similar to AF181721 Homo sapiens RU2S mRNA"
BASE COUNT 562 a 388 c 525 g 428 t
ORIGIN

Query Match 29.7% Score 409.8; DB 9; Length 2003;
Best Local Similarity 99.5% Pred. No. 2.4e-83;
Matches 411; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 554 ACTTGAGTTCTTGAAGGCTTCCTGGCTCCAGCAGTAAATGCCCGCTCTGGATCT 613
DB 413 AATTGAGTTCTTGAAGGCTTCCTGGCTCCAGCAGTAAATGCCCGCTCTGGATCT 354
QY 614 GGTCTAGTTCGGATTCCGTGGCCAGTCCGCGGGGTAGATGTTCTTGACGCCCA 673
DB 353 GGTCTAGTTCGGATTCCGTGGCCAGTCCGCGGGGTAGATGTTCTTGACGCCCA 294
QY 674 AGGTGCTGAACGCCCGCTCACTCTTCAGGAAGACTTCAAGCTGGACACCTTCT 733
DB 293 AGGTGCTGAACGCCCGCTCACTCTTCAGGAAGACTTCAAGCTGGACACCTTCT 234
QY 734 TCTCATGATGACGACGCGCGCGCTAGAGAGGGTCCCGCTTGGCGTACACAAGCA 793
DB 233 TCTCATGATGACGACGCGCGCGCTAGAGAGGGTCCCGCTTGGCGTACACAAGCA 174
QY 794 CGCTCTTACGACGCGGTGACAGAGTGGCTGGACCTGGCGCTCTGCGCTCATCTCC 853
DB 173 CGCTCTTACGACGCGGTGACAGAGTGGCTGGACCTGGCGCTCTGCGCTCATCTCC 114
QY 854 CCGCTGCGCGCGCTCAGCTCGCTGCTTCCGCTGGGAGGACCTCCGCTGTCACGCG 913
DB 113 CCGCTGCGCGCGCTCAGCTCGCTGCTTCCGCTGGGAGGACCTCCGCTGTCACGCG 54
QY 914 GCCTCAGCAGCAGCGCGGGGATGCTCTCTGAAACGAGAACTGA 966
DB 53 GCCTCAGCAGCAGCGCGGGATGCTCTCTGAAACGAGAACTGA 1

RESULT 8
AB032980/c 5548 bp mRNA linear PRI 10-MAY-2002
DEFINITION Homo sapiens mRNA for KIAA1154 protein, partial cds.
ACCESSION AB032980
VERSION AB032980.2 GI:20521781
KEYWORDS Homo sapiens adult male brain cDNA to mRNA, clone_lib:pBluescriptII
SOURCE SK plus clone:hh03679s1.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Hirose, M., Nagase, T., Ishikawa, K., Kikuno, R., Nomura, N. and Ohara, O.
TITLE Characterization of cDNA clones selected by the GeneMark analysis from size-fractionated cDNA libraries from human brain
JOURNAL DNA Res. 6 (5), 329-336 (1999)
MEDLINE 20039618
PUBMED 10574461
REFERENCE 2 (bases 1 to 6548)
AUTHORS Ohara, O., Nagase, T. and Kikuno, R.
TITLE Direct Submission
JOURNAL Submitted (04-OCT-1999) Osamu Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp, URL: http://www.kazusa.or.jp/huge/, Tel:+81-438-52-3913,
```

REFERENCE	1 (bases 1 to 19957)	RESULT 10
AUTHORS	Oliver, K.	AC125638
TITLE	Direct.Submission	LOCUS
JOURNAL	Submitted (29-JUN-2002)	
	Wellcome Trust Sanger Institute, Hinxton,	

**COMMENT**

FEATURES  
SOURCE

BASE COUNT	60102 a	43920 c	42965 g	52970.t
ORIGIN				

RESULT 10	AC125638	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
-----------	----------	-------	------------	-----------	---------	----------	--------	----------

REFERENCE  
AUTHORS

1. (bases 1 to 159103)

Adams, C., Ali-oshman, F. R., Allen, C., Alsbrooks, S. L., Amaratunge, H. C., Are, J. R., Ayele, M., Banks, T., Barbarella, J., Benton, J., Binaige, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowle, S., Brieva, M., Brown, E., Brown, M., Bryant, N. P., Buhay, C., Burch, P., Burkett, C., Burrell, K. L., Byrd, N. C., Carron, T. F., Carter, M., Cavazos, S. R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C. D., Cox, C., Coyle, M. D., Dathorne, S. R., David, R., Devila, M. L., Davis, C., Davy-Carroll, L., Dederich, D. A., Delaney, K. R., Delgado, O., Denn, A. L., Ding, Y., Dinh, H. H., Douthwaite, K. J., Draper, H., Dugan-Rocha, S., Durbin, K. J., Earnhart, C., Edgar, D., Edwards, C. C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J. H., Guevarra, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Haves, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B., Homs, J., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L. E., Jacobson, B., Jia, Y., Johnson, R., Jollivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratochvic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseg, H., Lozaro, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M. P., Meador, K., Mel, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuono, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Picken, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansley, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, O., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE  
AUTHORS

TITLE  
JOURNAL

COMMENT

Center: Baylor College of Medicine  
Center code: BCM  
web site: <http://www.hqsc.bcm.tmc.edu/>



Qy 730 TTCTTTCATGATGACGACGCGCGCCGCGGTAGAGGGTCCCGCTTGGGTACACA 789  
 Db 139836 CGGCGCGGAGACACGAACTTAGCCGCCACATAGAACTGTCGCCCTGGGTACACC 139895

Qy 790 AGCAGCTCTT 800  
 Db 139896 AAGATGGTCTT 139906

RESULT 11  
 AC010907/c  
 LOCUS AC010907 139357 bp DNA linear PRI 09-JAN-2002  
 DEFINITION Homo sapiens BAC clone RP11-568H24 from 2, complete sequence.  
 ACCESSION AC010907  
 VERSION AC010907.10 GI:15321567  
 KEYWORDS HTG.  
 SOURCE  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 139357)  
 AUTHORS Sulston, J.E. and Waterston, R.  
 TITLE Toward a complete human genome sequence  
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)  
 MEDLINE 99063792  
 PUBMED 9847074  
 REFERENCE 2 (bases 1 to 139357)  
 AUTHORS Du, H., Haakenson, W. and Dixon, R.  
 TITLE The sequence of Homo sapiens BAC clone RP11-568H24  
 JOURNAL Unpublished (2001)  
 REFERENCE 3 (bases 1 to 139357)  
 AUTHORS Waterston, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (25-SEP-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 REFERENCE 4 (bases 1 to 139357)  
 AUTHORS Waterston, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (28-AUG-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 REFERENCE 5 (bases 1 to 139357)  
 AUTHORS Waterston, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 COMMENT On Aug 28, 2001 this sequence version replaced gi:13399437.  
 ----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu/gsc>  
 Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)  
 ----- Summary Statistics  
 -----  
 -----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D.

McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Frengen, E., Taten, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)  
 VECTOR: pBACE3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RP11-178E6, 2000 bp overlap. Actual start of this clone is at base position 1 of RP11-568H24.

The sequence between 66093 to 66578 and 104506 to 104590 is covered only by PCR products from clone DNA. The sequence contains a dinucleotide (TC) run from 65513 to 65634 in which the exact length is unknown. The sequence contains a dinucleotide (TC) run from 104386 bp to 104631 bp in which the exact length is unknown. The sequence from base position 4458 to 6187 can not be guaranteed due to a tandem repeat.

#### FEATURES

	Location/Qualifiers
Source	1..139357
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="2"
	/map="2"
	/clone_lib="RPCI-11"
	383..409
repeat_region	/rpt_family="AT-rich"
	876..950
repeat_region	/rpt_family="Achobo"
	951..1248
repeat_region	/rpt_family="Alu"
	1249..1339
repeat_region	/rpt_family="Achobo"
	1340..1514
repeat_region	/rpt_family="MERL_type"
	1611..1965
repeat_region	/rpt_family="Achobo"
	1973..2279
repeat_region	/rpt_family="Alu"
	2324..2843
repeat_region	/rpt_family="L1"
	2844..3050
repeat_region	/rpt_family="MERL_type"
	2943..2958
misc_feature	/note="similar to Homo sapiens EST AI597790 (NID:94606838) tr92h02.xl"
	3051..3174
repeat_region	/rpt_family="L1"
	3175..3483
repeat_region	/rpt_family="Alu"
	3182..3206
misc_feature	/note="similar to Homo sapiens EST BGI78637 (NID:912685340)"
	3484..4227
repeat_region	/rpt_family="L1"
	4255..4330
repeat_region	/rpt_family="(T)n"
	4312..4987
repeat_region	/rpt_family="SVA"
	4571..6524
misc_feature	/note="CpG island (%GC=74.6, o/e=0.80, #CpGs=159)"
	4978..5152
repeat_region	/rpt_family="(CCGG)n"
	5163..5560
repeat_region	/rpt_family="SVA"

repeat_region	5249..5273	/rpt_family="GC-rich"	
repeat_region	5593..6486	/rpt_family="SVA"	
repeat_region	6443..6529	/rpt_family="GA-rich"	
repeat_region	6531..6732	/rpt_family="L1"	
repeat_region	6733..7033	/rpt_family="Alu"	
repeat_region	7034..7249	/rpt_family="L1"	
repeat_region	7250..7534	/rpt_family="Alu"	
repeat_region	7555..8490	/rpt_family="L1"	
repeat_region	8528..8726	/rpt_family="Achoho"	
misc_feature	8666..8669	/note="match to EST BF314952 (NID:g11263123)"	
misc_feature	9031..9360	/note="match to EST BF314952 (NID:g11263123)"	
misc_feature	9047..9360	/note="match to EST BE259985 (NID:g9130920)"	
misc_feature	9088..9089	/note="match to EST BF316717 (NID:g11265074)"	
misc_feature	9088..9069	/note="match to EST BI198782 (NID:g14653803)"	
misc_feature	9144..9147	/note="similar to Homo sapiens EST BF311850 (NID:g11259616)"	
misc_feature	9239..9360	/note="match to EST BI198831 (NID:g14653852)"	
misc_feature	9239..9360	/note="similar to Homo sapiens EST BI198308 (NID:g14653329)"	
misc_feature	9239..9360	/note="similar to Homo sapiens EST BI198502 (NID:g14653523)"	
misc_feature	9240..9360	/note="match to EST BF206235 (NID:g11099821)"	
misc_feature	9240..9360	/note="match to EST BF311988 (NID:g11259763)"	
misc_feature	9240..9360	/note="match to EST BI199896 (NID:g14654917)"	
misc_feature	9240..9360	/note="similar to Homo sapiens EST BF313026 (NID:g11260952)"	
misc_feature	9313..9360	/note="match to EST BF311981 (NID:g11259754)"	
misc_feature	9313..9360	/note="similar to Homo sapiens EST BF316085 (NID:g11264380)"	
misc_feature	9313..9360	/note="similar to Homo sapiens EST BI198293 (NID:g14653314)"	
misc_feature	9322..9360	/note="match to EST BF316717 (NID:g11265074)"	
misc_feature	9322..9360	/note="match to EST BI198782 (NID:g14653803)"	
misc_feature	9322..9360	/note="similar to Homo sapiens EST BF311850 (NID:g11259616)"	
misc_feature	9333..9360	/note="match to EST BE313758 (NID:g9134050)"	
misc_feature	9338..9360		
Query Match	6.28;	Score 85.4;	DB 9; Length 139357;
Best Local Similarity	59.18;	Pred. No. 2e-08;	
Matches 165; Conservative	0;	Mismatches 111;	Indels 3; Gaps 1;
Oy	526	ATGGGTGGCGGTGGGGAGACCGACTCTGAGTTCTTGAAGGCTTCTCGCCTCCA	585

Db 118260 AGGCGCGTGGGGGCGCGGCACATTACTCGAGCTCTTGAACGCTCGCGGCCGCC 118201

Oy 586 GCCAGGTAATTGCCCGCTCTGGATCTGGTCTAGCTTCCGGATTCGGTGGCCAGTCCGC 645

Db 118200 GCCAGGTACTTGGCGCGCTGCGAGCGGTCCAGCCCGACCGCGTGGCCCGCGTG 118141

Oy 646 GGGGTGTAGATGTTCTTGACGCGCCCAAGGTTGCTTGAACGCGCGCGGTACCTCCCTTC 705

Db 118140 GGCCTGAAGAGCGCGCGCACG---CCGAACGGGAGCTCCACCTGCTCGGTGAGTGTCTCC 118084

Oy 706 AGGAAGACTTGAAGCTGGACACCTTCTTCTCATGTGATGACGACGCGCGCGCTAG 765

Db 118083 AGCAGCGCTCGAAGTGGCGCGCGCGCGCGACACGACGAACTTCTTCCCGACGTAG 118024

Oy 766 AAGGGTCCCGTTCGCGTACACAGACGCTCTTCACG 804

Db 118023 AACGGTCCCGTTCGCGTACACGACGATGCTCTTGGCG 117985

RESULT 12

HSDJ622L5/c

LOCUS HSDJ622L5 115756 bp DNA linear PRI 22-NOV-2001

DEFINITION Human DNA sequence from clone RP4-622L5 on chromosome 1p34.2-36.11, complete sequence.

ACCESSION AL049795

VERSION AL049795.21 GI:17065925

KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 115756)

AUTHORS Frankland,J.

TITLE Direct Submission

JOURNAL Submitted (22-NOV-2001) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: hamquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Nov 25, 2001 this sequence version replaced gi:6010175. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C-elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chrl This sequence is the entire insert of clone RP4-622L5 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP4-622L5 is from the library RPCI-4 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm VECTOR: pCYPAC2.

FEATURES

source

1. .115756

Location/Qualifiers

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="1"

/map="p34.2-36.11"

/clone="RP4-622L5"

/clone\_lib="RPCI-4"

```
repeat_region 1645. .1810
/note="TIGER1 repeat: matches 2259. .2417 of consensus"
repeat_region 2123. .2437
/note="TIGER1 repeat: matches 1957. .2259 of consensus"
repeat_region 2726. .2782
/note="TIGER1 repeat: matches 1904. .1957 of consensus"
repeat_region 3067. .3486
/note="TIGER1 repeat: matches 1501. .1904 of consensus"
repeat_region 3781. .4180
/note="TIGER1 repeat: matches 1117. .1501 of consensus"
repeat_region 5033. .5082
/note="TIGER1 repeat: matches 1090. .1117 of consensus"
repeat_region 5372. .5501
/note="TIGER1 repeat: matches 966. .1090 of consensus"
repeat_region 5628. .5704
/note="TIGER1 repeat: matches 893. .966 of consensus"
repeat_region 5991. .6492
/note="TIGER1 repeat: matches 411. .893 of consensus"
repeat_region 6800. .6835
/note="TIGER1 repeat: matches 378. .411 of consensus"
repeat_region 6968. .7022
/note="TIGER1 repeat: matches 1. .378 of consensus"
repeat_region 8281. .8358
/note="3 copies 26 mer 79% conserved"
repeat_region 8301. .8360
/note="30 copies 2 mer ta 81% conserved"
repeat_region 17674. .17777
/note="4 copies 26 mer 77% conserved"
repeat_region 17685. .17784
/note="50 copies 2 mer at 82% conserved"
misc_feature 19385. .19542
/note="match: GSS: Em:AQ306911"
repeat_region 19551. .19574
/note="12 copies 2 mer tt 100% conserved"
repeat_region 20865. .20922
/note="29 copies 2 mer at 70% conserved"
repeat_region 20934. .21043
/note="55 copies 2 mer at 70% conserved"
repeat_region 20940. .21043
/note="4 copies 26 mer 74% conserved"
repeat_region 21168. .21234
/note="tRNA-Asp-GAC repeat: matches 1. .75 of consensus"
misc_feature 28263. .28691
/note="match: GSS: Em:AQ317905"
misc_feature complement(29964. .30447)
/note="match: GSS: Em:AQ732256"
misc_feature 30465. .30718
/note="match: STS: Em:G58909
match: GSS: Em:AQ376964"
repeat_region 38923. .39446
/note="L1MBC repeat: matches 1546. .2095 of consensus"
repeat_region 39761. .39882
/note="L1MBC repeat: matches 1429. .1546 of consensus"
misc_feature complement(43254. .43726)
/note="match: GSS: Em:AQ146371"
repeat_region 45555. .46050
/note="MER44C repeat: matches 1. .728 of consensus"
misc_feature 46055. .46468
/note="match: GSS: Em:AQ404362"
repeat_region 48914. .48982
/note="MER21B repeat: matches 718. .790 of consensus"
repeat_region 48985. .49173
/note="MER21B repeat: matches 77. .263 of consensus"
misc_feature 49285. .49651
/note="match: GSS: Em:AQ761532"
misc_feature 52462. .53177
/note="match: GSS: Em:AQ404082"
misc_feature 54716. .54865
/note="match: GSS: Em:AQ680771"
repeat_region 60551. .60614
/note="32 copies 2 mer tt 70% conserved"
misc_feature complement(62219. .62382)
/note="match: STS: Em:G20585"
```

```
misc_feature complement(66030. .66267)
/note="match: GSS: Em:AQ514287"
misc_feature complement(66041. .66267)
/note="match: GSS: Em:AQ803288"
misc_feature 66066. .66255
/note="match: GSS: Em:B69239"
misc_feature 68882. .69541
/note="CpG island"
/evidence=not_experimental
repeat_region 71427. .71448
/note="11 copies 2 mer tt 100% conserved"
misc_feature 73527. .74249
/note="match: GSS: Em:AQ892680"
misc_feature 74314. .74926
/note="match: GSS: Em:AQ441232"
misc_feature 74532. .74981
/note="match: GSS: Em:AQ812740"
misc_feature 74626. .75008
/note="match: GSS: Em:AQ091614"
misc_feature complement(74682. .75020)
/note="match: STS: Em:L29744"
misc_feature complement(74683. .74951)
/note="match: STS: Em:L29729 Em:L29763"
misc_feature 74717. .74867
/note="match: STS: Em:G15744"
misc_feature 82892. .83089
/note="match: STS: Em:G09447"
misc_feature 83391. .83481
/note="match: GSS: Em:A2310334"
repeat_region 85284. .85337
/note="27 copies 2 mer tt 88% conserved"
misc_feature 85339. .85718
/note="match: STS: Em:G28385"
misc_feature complement(87250. .87612)
/note="match: STS: Em:G06968"
misc_feature complement(87455. .87579)
/note="match: STS: Em:G26511"
misc_feature complement(87851. .88313)
/note="match: GSS: Em:AQ536603"
repeat_region 89654. .89685
/note="16 copies 2 mer tc 87% conserved"
misc_feature 91856. .92386
/note="match: GSS: Em:B85187"
repeat_region 92101. .92152
/note="26 copies 2 mer tt 71% conserved"
misc_feature 94297. .95612
/note="CpG island"
/evidence=not_experimental
misc_feature complement(98493. .98894)
/note="match: GSS: Em:AQ207463"
misc_feature 103719. .104155
/note="match: GSS: Em:AQ543924"
repeat_region 104406. .104437
/note="MER57B repeat: matches 303. .334 of consensus"
repeat_region 107785. .107841
/note="L1MB1 repeat: matches 6111. .6171 of consensus"
misc_feature 111249. .111767
/note="CpG island"
/evidence=not_experimental
misc_feature 113362. .113450
/note="match: STS: Em:G16146
match: GSS: Em:AQ806873"
BASE COUNT 29576 a 26665 c 27977 g 31538 t
ORIGIN
```

Query Match 5.5%; Score 75.8; DB 9; Length 115756;  
Best Local Similarity 58.2%; Pred.No.3e-06;  
Matches 153; Conservative 0; Mismatches 107; Indels 3; Gaps 1;

QY 547 CCGACTCACTTGAGTTTCTTGAAGGCTTCTGGCTCCAGCCACGTAATTCGCCCGCTC 606  
DB 98696 CCACACTTGAGCTTGGAATCGTTCAATCCAGGCCACACTACTGCCCTCTGTTTC 98637

Qy 607 TGGATCTGCTAGTTCCTGGATTGGCGGAGTCCCGGGGTGTAGATGTTCTGACG 666  
Db 98636 TTCAAGTTCGCGAGTGTGTCAGAGGTGGCGATGACAAAGGTGTGTAGAGGCGACGCACA 98577  
Qy 667 GCCCAAGAGGTGCTGAACCGCCGCGTCACTTTCAGAGAGACTTCGAAGCTGCAC 726  
Db 98576 GCC---AGTGGGGCTGCACAGCTGATGTCACTTCGCGAGAGAGGCTCCATGTTGGGG 98520  
Qy 727 ACCTTCTTCATGATGACGACGCGCGCCCGCGTGAAGGGTCCCGCTTGGGTTAC 786  
Db 98519 AAGCGGCTTCAGTCAACACAGCTGGGAGCGCTGGGAAGATGGTCCCATTCGGTAC 98460  
Qy 787 ACAGCAGCTTCACACAGGG 809  
Db 98459 ACCACTACCTCTTGGCTGCTGG 98437

RESULT 13  
HSDJ675E8/c  
LOCUS HSDJ675E8 322840 bp DNA linear HTG 17-JUL-2002  
DEFINITION Homo sapiens chromosome 1 clone RP4-675E8 map p34.1-35.3, \*\*\*  
SEQUENCING IN PROGRESS \*\*\*, 45 unordered pieces.  
ACCESSION AL121991  
VERSION AL121991.41 GI:21912433  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_ACTIVEFIN; HTGS\_DRAFT; HTGS\_FULLTOP.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 322840)  
Matthews.N.  
Direct Submission  
Submitted (15-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Jul 19, 2002 this sequence version replaced gi:21535828.  
COMMENT ----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: humquerry@sanger.ac.uk  
----- Project Information  
Center project name: du675E8  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: plasmid; 108752; 100% of reads  
Chemistry: Dye-terminator ABI; 0% of reads  
Chemistry: Dye-terminator; 8% of reads  
Chemistry: Dye-terminator ET-amersham; 4% of reads Chemistry:  
Dye-terminator Big Dye; 86% of reads  
Consensus quality: 309409 bases at least Q40  
Consensus quality: 313127 bases at least Q30  
Consensus quality: 315404 bases at least Q20  
Insert size: 318440; sum-of-contigs  
Insert size: 128138; 23.4% error; agarose-fp  
Quality coverage: 6.9% in Q20 bases; sum-of-contigs Quality  
coverage: 22.25x in Q20 bases; agarose-fp  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 45 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 54147: contig of 54147 bp in length  
\* 54148 54247: gap of 100 bp  
\* 54248 71025: contig of 16778 bp in length  
\* 71026 71125: gap of 100 bp  
\* 71126 75812: contig of 4687 bp in length  
\* 75813 75912: gap of 100 bp  
\* 75913 116848: contig of 40936 bp in length

\* 116849 116948: gap of 100 bp  
\* 116949 118963: contig of 2015 bp in length  
\* 118964 119063: gap of 100 bp  
\* 119064 125611: contig of 6548 bp in length  
\* 125612 125711: gap of 100 bp  
\* 125712 164112: contig of 38401 bp in length  
\* 164113 164212: gap of 100 bp  
\* 164213 180038: contig of 15826 bp in length  
\* 180039 180138: gap of 100 bp  
\* 180139 184418: contig of 4280 bp in length  
\* 184419 184518: gap of 100 bp  
\* 184519 187132: contig of 2614 bp in length  
\* 187133 187232: gap of 100 bp  
\* 187233 190323: contig of 3091 bp in length  
\* 190324 190423: gap of 100 bp  
\* 190424 192746: contig of 2323 bp in length  
\* 192747 192846: gap of 100 bp  
\* 192847 194893: contig of 2047 bp in length  
\* 194894 194993: gap of 100 bp  
\* 194994 197169: contig of 2176 bp in length  
\* 197170 197269: gap of 100 bp  
\* 197270 199729: contig of 2460 bp in length  
\* 199730 199829: gap of 100 bp  
\* 199830 202634: contig of 2805 bp in length  
\* 202635 202734: gap of 100 bp  
\* 202735 205712: contig of 2978 bp in length  
\* 205713 205812: gap of 100 bp  
\* 205813 208288: contig of 2476 bp in length  
\* 208289 208388: gap of 100 bp  
\* 208389 213338: contig of 4950 bp in length  
\* 213339 213438: gap of 100 bp  
\* 213439 216918: contig of 3480 bp in length  
\* 216919 217018: gap of 100 bp  
\* 217019 220931: contig of 3913 bp in length  
\* 220932 221031: gap of 100 bp  
\* 221032 224098: contig of 3067 bp in length  
\* 224099 224198: gap of 100 bp  
\* 224199 226367: contig of 2169 bp in length  
\* 226368 226467: gap of 100 bp  
\* 226468 229613: contig of 3146 bp in length  
\* 229614 229713: gap of 100 bp  
\* 229714 232104: contig of 2391 bp in length  
\* 232105 232204: gap of 100 bp  
\* 232205 265235: contig of 33031 bp in length  
\* 265236 265335: gap of 100 bp  
\* 265336 268023: contig of 2688 bp in length  
\* 268024 268123: gap of 100 bp  
\* 268124 270219: contig of 2096 bp in length  
\* 270220 270319: gap of 100 bp  
\* 270320 278019: contig of 7700 bp in length  
\* 278020 278119: gap of 100 bp  
\* 278120 280129: contig of 2010 bp in length  
\* 280130 280229: gap of 100 bp  
\* 280230 282545: contig of 2316 bp in length  
\* 282546 282645: gap of 100 bp  
\* 282646 285394: contig of 2749 bp in length  
\* 285395 285494: gap of 100 bp  
\* 285495 287702: contig of 2208 bp in length  
\* 287703 287802: gap of 100 bp  
\* 287803 289871: contig of 2069 bp in length  
\* 289872 289971: gap of 100 bp  
\* 289972 293367: contig of 3396 bp in length  
\* 293368 293467: gap of 100 bp  
\* 293468 295482: contig of 2015 bp in length  
\* 295483 295582: gap of 100 bp  
\* 295583 299991: contig of 4409 bp in length  
\* 299992 300091: gap of 100 bp  
\* 300092 302207: contig of 2116 bp in length  
\* 302208 302307: gap of 100 bp  
\* 302308 304990: contig of 2683 bp in length  
\* 304991 305090: gap of 100 bp  
\* 305091 307235: contig of 2245 bp in length  
\* 307236 307435: gap of 100 bp

307436	309641: contig of 2206 bp in length
309642	309741: gap of 100 bp
309742	315531: contig of 5790 bp in length
315531	315631: gap of 100 bp
315632	318321: contig of 2690 bp in length
315632	318421: gap of 100 bp
318322	318421: gap of 100 bp
318422	320459: contig of 2048 bp in length
320459	320569: gap of 100 bp
320470	322840: contig of 2271 bp in length.
320570	

## FEATURES

```
location="/quadrialeis
source
1..322840
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="p34.1-35.3"
/clone="RP4-675P8"
/clone_lib="RPCI-4"
1..34147
/note="assembly_fragment:11504
fragment_chain:1"
54248..71025
/note="assembly_fragment:09814
fragment_chain:1"
71126..75812
/note="assembly_fragment:11398
fragment_chain:1"
75913..116848
/note="assembly_fragment:03358
fragment_chain:1"
116949..118963
/note="assembly_fragment:10591
fragment_chain:2"
119064..125611
/note="assembly_fragment:09365
fragment_chain:2"
125712..164112
/note="assembly_fragment:11336
fragment_chain:3"
164213..180038
/note="assembly_fragment:02790
fragment_chain:3"
180139..184418
/note="assembly_fragment:00167"
184519..187132
/note="assembly_fragment:00622"
187233..190323
/note="assembly_fragment:01142"
190424..192746
/note="assembly_fragment:01318"
192847..194893
/note="assembly_fragment:01614"
194994..197169
/note="assembly_fragment:01627"
197270..199729
/note="assembly_fragment:01811"
199830..202634
/note="assembly_fragment:01839"
202735..205712
/note="assembly_fragment:02187"
205813..208288
/note="assembly_fragment:02212"
208389..213338
/note="assembly_fragment:02545"
213439..216918
/note="assembly_fragment:02573"
217019..220931
/note="assembly_fragment:02907"
221032..224098
/note="assembly_fragment:04499"
224199..226367
/note="assembly_fragment:05011"
226468..229613
/note="assembly_fragment:05130"
```



RESULT 15

AB047038/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

AB047038 3295 bp mRNA linear INV 11-JAN-2002

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

Db 289 ACCAAGTTTAGAAATCCATCGAAATCGCATATGTATCTTCTGTGACGACGAACCTTCGT 230

Qy 757 CCCCGGTAGAGGGGTCCCCGTTGGGTACACAGCAGCGCTCTTCAC 803

Db 229 CCGGGGTAGTACTGATCCCCGTTCCACGTAAACGTGAACATTTTTCGC 183

Search completed: November 26, 2002, 05:30:43  
Job time : 5060 secs

AB047038 3295 bp mRNA linear INV 11-JAN-2002

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

BASE COUNT 1153 a 651 c 680 g 811 t  
ORIGIN

Query Match 4.9%; Score 67.8; DB 3; Length 3295;  
Best Local Similarity 52.3%; Pred. No. 9.8e-05;  
Matches 150; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

Qy 517 GGCACCTAATGGTGGGGGTGGGGAGACCGACTACTTGAGTTCTTGAAGGCTTCC 576

Db 469 GGCATCGGTTTCTTGGAGTGGAGTGTGTTTGAATGAAGTCTTCTGAATCTTCT 410

Qy 577 TGGCCTCCAGCCAGTAATCCGCCCTCTGGATCTGGTCTCCGGATTCGGTGG 636

Db 409 GCTCCGGCGGCAACAATACCATGTTATGTTAATTCGGAAGTTCGTTAACCCGATGT 350

Qy 637 CCAGTCCCGGGGTGTAGATGTTCTTACGCGCCCAAGAGGTGCTGAACCCCGCGTC 696

Db 349 CCCCGTTGGTGTGTAGATGTTCCGCACAGCGCCGAAAGACGTTTCAGGCCGCTTGA 290

Qy 697 ACCTCTTCAGGAAGATTGGAAGCTGGACACCTTCTTCATGATGACGACGCGCGC 756

Db 697 ACCTCTTCAGGAAGATTGGAAGCTGGACACCTTCTTCATGATGACGACGCGCGC 756

Qy 697 ACCTCTTCAGGAAGATTGGAAGCTGGACACCTTCTTCATGATGACGACGCGCGC 756

Db 697 ACCTCTTCAGGAAGATTGGAAGCTGGACACCTTCTTCATGATGACGACGCGCGC 756

Qy 697 ACCTCTTCAGGAAGATTGGAAGCTGGACACCTTCTTCATGATGACGACGCGCGC 756

Db 697 ACCTCTTCAGGAAGATTGGAAGCTGGACACCTTCTTCATGATGACGACGCGCGC 756

Qy 697 ACCTCTTCAGGAAGATTGGAAGCTGGACACCTTCTTCATGATGACGACGCGCGC 756

Db 697 ACCTCTTCAGGAAGATTGGAAGCTGGACACCTTCTTCATGATGACGACGCGCGC 756

Qy 697 ACCTCTTCAGGAAGATTGGAAGCTGGACACCTTCTTCATGATGACGACGCGCGC 756

Db 697 ACCTCTTCAGGAAGATTGGAAGCTGGACACCTTCTTCATGATGACGACGCGCGC 756

Qy 697 ACCTCTTCAGGAAGATTGGAAGCTGGACACCTTCTTCATGATGACGACGCGCGC 756

Db 697 ACCTCTTCAGGAAGATTGGAAGCTGGACACCTTCTTCATGATGACGACGCGCGC 756

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 25, 2002, 09:07:35 ; Search time 60 Seconds

(without alignments)  
7063.788 Million cell updates/sec

Title: US-09-674-593-1

Perfect score: 1382

Sequence: 1 cattatgctaacagcataaa.....taagcaattaagtctctgg 1382

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued\_Patents\_NA:\*
- 1: /cgn2\_6/ptodata/1/ina/5A-COMB.seq.\*
- 2: /cgn2\_6/ptodata/1/ina/5B-COMB.seq.\*
- 3: /cgn2\_6/ptodata/1/ina/6A-COMB.seq.\*
- 4: /cgn2\_6/ptodata/1/ina/6B-COMB.seq.\*
- 5: /cgn2\_6/ptodata/1/ina/PCTUS-COMB.seq.\*
- 6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
- score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
C 1	46.8	3.4	7218	1	US-08-232-463-14
C 2	44.8	3.2	1212	4	US-09-182-145-34
C 3	44.8	3.2	1212	4	US-09-182-145-35
C 4	43.4	3.1	8438	1	US-07-945-283-1
C 5	43	3.1	1519	1	US-07-971-759-19
C 6	42.2	3.1	1441	4	US-08-821-994-63
C 7	42.2	3.1	3238	4	US-08-123-934A-5
C 8	42.2	3.1	3238	5	PCT-US94-10080-5
C 9	40.2	2.9	2581	4	US-09-363-708-1
C 10	39.8	2.9	4403765	4	US-09-103-840A-2
C 11	39.4	2.9	688	6	5498694-3
C 12	39.4	2.9	2648	4	US-09-417-455-6
C 13	39.4	2.9	2648	4	US-09-348-942-6
C 14	39.4	2.9	2648	4	US-09-457-626-6
C 15	39.2	2.8	732	4	US-09-149-476-66
C 16	39	2.8	746	4	US-09-013-810-1
C 17	39	2.8	1176	4	US-09-372-422A-25
C 18	38.8	2.8	1931	3	US-09-019-942-2
C 19	38.8	2.8	1931	4	US-09-099-041A-1
C 20	38.8	2.8	1931	4	US-09-245-281-1
C 21	38.8	2.8	1931	4	US-09-470-271-2
C 22	38.8	2.8	1931	4	US-09-207-359B-1
C 23	38.8	2.8	80246	4	US-09-078-294-4
C 24	38.8	2.8	80595	4	US-09-078-294-3
C 25	38.6	2.8	1114	4	US-09-152-060-41
C 26	38.4	2.8	1740	4	US-09-125-642C-2
C 27	38.4	2.8	1742	4	US-09-125-642C-13

Sequence 1, Appli	28	38.4	2.8	2239	4	US-09-196-390-1
Sequence 8, Appli	C 29	38.4	2.8	5515	4	US-09-125-642C-8
Sequence 12, Appli	C 30	38.4	2.8	5519	4	US-09-125-642C-12
Sequence 16, Appli	C 31	38.4	2.8	9468	1	US-08-325-547-10
Sequence 2, Appli	C 32	38.2	2.8	1738	2	US-08-379-482A-2
Sequence 11, Appli	C 33	38	2.7	948	2	US-08-924-759-11
Sequence 11, Appli	C 34	38	2.7	948	3	US-09-248-335-11
Sequence 4, Appli	C 35	38	2.7	1046	1	US-08-361-467B-4
Sequence 4, Appli	C 36	38	2.7	1046	1	US-08-484-332C-4
Sequence 3, Appli	C 37	38	2.7	2760	1	US-08-101-593-3
Sequence 3, Appli	C 38	37.6	2.7	1364	1	US-08-265-087-3
Sequence 3, Appli	C 39	37.6	2.7	1364	1	US-08-621-493-3
Sequence 3, Appli	C 40	37.6	2.7	1364	2	US-08-965-688-3
Sequence 3, Appli	C 41	37.6	2.7	1364	4	US-09-260-173-3
Sequence 7, Appli	C 42	37.6	2.7	1374	2	US-08-758-621-7
Sequence 7, Appli	C 43	37.6	2.7	1374	4	US-09-107-858-7
Sequence 13, Appli	C 44	37.6	2.7	2448	4	US-08-487-596-13
Sequence 9, Appli	C 45	37.6	2.7	2450	2	US-08-466-589-9

ALIGNMENTS

RESULT 1  
US-08-232-463-14/C  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pfz9pt-fls  
US-08-232-463-14

Query Match 3.4%; Score 46.8; DB 1; Length 7218;

```
Best Local Similarity 2.2%; Pred. No. 0.018;
Matches 6; Conservative 169; Mismatches 101; Indels 0; Gaps 0;

Qy 10 AACAGCATAACATGACGAGGGTGGGAGCGGTCCACAAAAGTGAGTGTCTCAATTCTA 69
Db 1334 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1275
Qy 70 CTGGAATGAAGGTGAATTAATTAACAGTACGGGAATGCAGAGCAATTTCTCTCT 129
Db 1274 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1215
Qy 130 CTGGTGACAATAGTGTCCAACTGCGAAGTATTTTAAAGATGTTTATTAAATTA 189
Db 1214 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1155
Qy 190 AAAGATGATTCACAGGAAAGAAATAAGGAAAGGAAAGAAAGAAAGAAAGAAAG 249
Db 1154 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1095
Qy 250 CGCAAAGTATCAGTTGGTCACTAACCTTTGCAAG 285
Db 1094 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1059

RESULT 2
US-09-182-145-34
; Sequence 34, Application US/09182145B
; Patent No. 6387657
; GENERAL INFORMATION:
; APPLICANT: Botstein, David A.
; APPLICANT: Cohen, Robert
; APPLICANT: Goddard, Audrey
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Lawrence, David A.
; APPLICANT: Levine, Arnold J.
; APPLICANT: Pennica, Diane
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: WISP POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: P1176R2
; CURRENT APPLICATION NUMBER: US/09/182,145B
; CURRENT FILING DATE: 1998-10-29
; EARLIER APPLICATION NUMBER: US 60/063,704
; EARLIER FILING DATE: 1997-10-29
; EARLIER APPLICATION NUMBER: US 60/073,612
; EARLIER FILING DATE: 1998-02-04
; EARLIER APPLICATION NUMBER: US 60/081,695
; EARLIER FILING DATE: 1998-04-14
; NUMBER OF SEQ ID NOS: 156
; SEQ ID NO 34
; LENGTH: 1212
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-182-145-34

Query Match 3.2%; Score 44.8; DB 4; Length 1212;
Best Local Similarity 55.0%; Pred. No. 0.024;
Matches 88; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

Qy 96 AAACAGTACGGGAATGCAGAGCAATTTCTCTCTGTGACAATATAGTGTCCAACT 155
Db 181 AAATGTCAGAGAACCTGGAGATATATTTTCTGAGCTCAAGATTTCTGAAAACCAAGCAA 122
Qy 156 TGGAGTGTATTTTAAAGATGTTTATTTAAATTAAGAGTGGATTTCCAGGAAAAAAA 215
Db 121 TGGGGGAAAAGTTAGTCAATCTGTCATATATAAAAAATTAGTCAGTAAAAAAA 62
Qy 216 ATAAGGAAAAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 255
Db 61 AAAAAAAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 22

RESULT 4
US-07-945-283-1/c
; Sequence 1, Application US/07945283
; Patent No. 5352596
; GENERAL INFORMATION:
; APPLICANT: Cheung, Andrew K.
; APPLICANT: Wesley, Ronald D.
; TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants
; TITLE OF INVENTION: Involving The EP0 and LLT Genes
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis P. Ribando
; STREET: 1815 No. 5352596th University Street
; CITY: Peoria
; STATE: IL
; COUNTRY: USA
; ZIP: 61604
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/945,283
```

```
RESULT 3
US-09-182-145-35/c
; Sequence 35, Application US/09182145B
; Patent No. 6387657
; GENERAL INFORMATION:
; APPLICANT: Botstein, David A.
; APPLICANT: Cohen, Robert
; APPLICANT: Goddard, Audrey
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Lawrence, David A.
; APPLICANT: Levine, Arnold J.
; APPLICANT: Pennica, Diane
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: WISP POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: P1176R2
; CURRENT APPLICATION NUMBER: US/09/182,145B
; CURRENT FILING DATE: 1998-10-29
; EARLIER APPLICATION NUMBER: US 60/063,704
; EARLIER FILING DATE: 1997-10-29
; EARLIER APPLICATION NUMBER: US 60/073,612
; EARLIER FILING DATE: 1998-02-04
; EARLIER APPLICATION NUMBER: US 60/081,695
; EARLIER FILING DATE: 1998-04-14
; NUMBER OF SEQ ID NOS: 156
; SEQ ID NO 35
; LENGTH: 1212
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-182-145-35

Query Match 3.2%; Score 44.8; DB 4; Length 1212;
Best Local Similarity 55.0%; Pred. No. 0.024;
Matches 88; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

Qy 96 AAACAGTACGGGAATGCAGAGCAATTTCTCTCTGTGACAATATAGTGTCCAACT 155
Db 181 AAATGTCAGAGAACCTGGAGATATATTTTCTGAGCTCAAGATTTCTGAAAACCAAGCAA 122
Qy 156 TGGAGTGTATTTTAAAGATGTTTATTTAAATTAAGAGTGGATTTCCAGGAAAAAAA 215
Db 121 TGGGGGAAAAGTTAGTCAATCTGTCATATATAAAAAATTAGTCAGTAAAAAAA 62
Qy 216 ATAAGGAAAAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 255
Db 61 AAAAAAAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 22

RESULT 4
US-07-945-283-1/c
; Sequence 1, Application US/07945283
; Patent No. 5352596
; GENERAL INFORMATION:
; APPLICANT: Cheung, Andrew K.
; APPLICANT: Wesley, Ronald D.
; TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants
; TITLE OF INVENTION: Involving The EP0 and LLT Genes
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis P. Ribando
; STREET: 1815 No. 5352596th University Street
; CITY: Peoria
; STATE: IL
; COUNTRY: USA
; ZIP: 61604
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/945,283
```

```
; FILING DATE: 19920911
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Ribando, Curtis P
; REGISTRATION NUMBER: 27976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 309-685-4011 ext.513
; TELEFAX: 309-685-4128
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8438 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Pseudorabies virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 622..6495
; FEATURE:
; NAME/KEY: variation
; LOCATION: replace(1099, "g")
; FEATURE:
; NAME/KEY: variation
; LOCATION: replace(1267, "c")
; FEATURE:
; NAME/KEY: variation
; LOCATION: replace(1381, "c")
; FEATURE:
; NAME/KEY: variation
; LOCATION: replace(1566, "c")
; FEATURE:
; NAME/KEY: variation
; LOCATION: replace(7010, "g")
; US-07-945-283-1

Query Match 3.1% Score 43.4; DB 1; Length 8438;
Best Local Similarity 57.9%; Pred. No. 0.16; Mismatches 56; Indels 0; Gaps 0;
Matches 77; Conservative 0;

Qy 790 AGCAGCTCTTCAGACGGCTGAGACAGGTGGCTGGACCTGGCGCTGCTCGCTCATC 849
|||||
Db 6349 AGCACTCGACGCGCAGAGCGCGCGCGCGCGCGCGCGCGCTCGTCCCTCGC 6290
|||||

Qy 850 TTCCCGCTGCGCGCGCTCAGCTCGCTTCTCGCTGCGGAGGACCTCCGCTGTC 909
|||||
Db 6289 AGCCTCCCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6230
|||||

Qy 910 AGCGGCTCACCG 922
|||||
Db 6229 GCGCGCGCGCGCG 6217

RESULT 5
US-07-971-759-19/c
; Sequence 19, Application US/07971759
; Patent No. 5573943
; GENERAL INFORMATION:
; APPLICANT: Saul, Alan J.
; APPLICANT: Cooper, Juan A.
; APPLICANT: Irving, David O.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF RHOPTRY
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
; STREET: 2100 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
```

```
; ZIP: 20037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/971.759
; FILING DATE: 01-FEB-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU91/00338
; FILING DATE: 01-AUG-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PK 1525 90
; FILING DATE: 02-AUG-1990
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1519 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 61..1254
; US-07-971-759-19

Query Match 3.1% Score 43; DB 1; Length 1519;
Best Local Similarity 54.9%; Pred. No. 0.082;
Matches 107; Conservative 0; Mismatches 85; Indels 3; Gaps 1;

Qy 150 AACACTTGGAGTGAATTTTAAAGATGTTTATTTTAAATTTAAAGATGATGATTTTCAAGGA 209
|||||
Db 1449 AATATTTAAATGTCAACTTAAAGAAATATTTTAAACAAACAAACAAACAAACAAACAAACAA 1390
|||||

Qy 210 AAAAAATTAAGAAAGAAAGAAAGAAACACT---GAACAGAAACAAACAAACAAACAAACAAACAA 266
|||||
Db 1389 AATAAATAACAAACAAAGAAAGTAAATAATTAATAAATAAACAACAAACAAACAAACAAACATATT 1330
|||||

Qy 267 GGTCACTAACCTTTTCAAGGATACCTTTTATTTTCTTTAAAGATTCCTGTTGTTTATACA 326
|||||
Db 1329 TATCAGAAACCTTTAATATATAAATGTAATTTTGTAAATAGTTAATGTTTTTTTATAA 1270
|||||

Qy 327 CAGATTTTAAAGTTTA 341
|||||
Db 1369 ATGTATTTTATTTTA 1255

RESULT 6
US-08-821-994-63
; Sequence 63, Application US/08821994A
; Patent No. 6228643
; GENERAL INFORMATION:
; APPLICANT: Greenland, Andrew J
; APPLICANT: Thomas, Didier RP
; APPLICANT: Jepson, Ian
; TITLE OF INVENTION: Promoters
; FILE REFERENCE: PPD 50108
; CURRENT APPLICATION NUMBER: US/08/821.994A
; CURRENT FILING DATE: 1997-03-22
; EARLIER APPLICATION NUMBER: PCT/GB97/00729
; EARLIER FILING DATE: 1997-03-18
; EARLIER APPLICATION NUMBER: GB 9606062.9
; EARLIER FILING DATE: 1996-03-22
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 63
; LENGTH: 1441
```









```
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 2648
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-457-626-6

Query Match      2.9%   Score 39.4;   DB 4;   Length 2648;
Best Local Similarity 52.1%   Pred. No. 1;
Matches 88;   Conservative 0;   Mismatches 81;   Indels 0;   Gaps 0;

Oy 88 AATAATTAAACAGTACGGGAATGACAGACCAATTTCTCTCTGGTGACAAATAGTGT 147
      ||| || || || || || || || || || || || || || || || || || ||
Db 2479 ATTAGTGAACCTAGAACGACAGACTCTGTGAGATAATCGATGTTTGTGTTTAAAGTTG 2538
      ||| || || || || || || || || || || || || || || || || || ||

Oy 148 CCAACACTTGAAGTGATTTTAAAGATGTTTATTTAAATTAAGAGTGCATTTCCAAG 207
      || || || || || || || || || || || || || || || || || || ||
Db 2539 CTCAGTTTGGTCTAACTGTTGTATGCGACCAATAGATAAATAATATGCGAGAGAAAGAGAAA 2598
      |||| || || || || || || || || || || || || || || || || || ||

Oy 208 GAAAAAATAAGGAAGGAAAGAAAAAACTGAACAGAAAAACGCAAAA 256
      |||| || || || || || || || || || || || || || || || || || ||
Db 2599 AAAAAAATAAGGAAGGAAAGAAAAAACTGAACAGAAAAACGCAAAA 2647
      |||| || || || || || || || || || || || || || || || || || ||

RESULT 15
US-09-149-476-66
; Sequence 66, Application US/09149476
; Patent No. 6420526
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: P2002p1
; CURRENT APPLICATION NUMBER: US/09/149,476
; CURRENT FILING DATE: 1998-09-08
; EARLIER APPLICATION NUMBER: PCT/JUS98/04493
; EARLIER FILING DATE: 1998-03-06
; EARLIER APPLICATION NUMBER: 60/040,162
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,333
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/038,621
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,626
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,334
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,336
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,163
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/047,600
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,615
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,597
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,502
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,633
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,583
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,617
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,618
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,503
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,592
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,581
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,584
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,500
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,587
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,492
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,598
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,613
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,582
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,596
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,612
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,632
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,601
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,580
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,568
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,314
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,569
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,311
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,671
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,674
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,669
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,312
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,313
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,672
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,315
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/048,974
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/056,886
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,877
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,889
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,893
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,630
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,878
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,662
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,872
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,882
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,637
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,903
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,888
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,879
```



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 25, 2002, 16:00:01 ; Search time 2051 seconds  
(without alignments)  
10912.805 Million cell updates/sec

Title: US-09-674-593-1  
Perfect score: 1382  
Sequence: 1 cattatgctaagcagcataaa.....taagcaattaagtctctgg 1382

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues  
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	776.2	56.2	879	14	BQ953400
2	730.2	52.8	1045	12	BG749180
3	665.6	48.2	845	14	BQ948660
C 4	609.6	44.1	898	14	BQ650550
C 5	534.4	38.7	963	14	BQ650598
C 6	470	34.0	486	14	BM833169

C 7	434.6	31.4	994	14	BQ650059
C 8	381.8	28.4	984	14	BM923267
C 9	288.4	20.9	929	14	BQ644671
C 10	288.4	20.9	939	14	BQ644474
C 11	288.4	20.9	1019	14	BM927835
C 12	252.8	18.3	477	14	BQ556967
C 13	252.8	18.3	634	10	BB660730
C 14	250.8	18.1	364	12	BF364659
C 15	244.4	17.7	246	9	AA863443
C 16	239.6	17.3	1048	14	BQ715572
C 17	232.8	16.8	923	12	BG751895
C 18	196.4	14.2	337	10	AW293816
C 19	192.2	13.9	835	12	BG750306
C 20	166.6	12.1	360	12	BF876266
C 21	145	10.5	683	17	AZ387429
C 22	122.8	8.9	525	13	BI841150
C 23	118.2	8.6	909	14	BQ731080
C 24	78.4	5.7	690	13	BQ072555
C 25	77.4	5.6	429	10	AW871951
C 26	75	5.4	665	14	BP027237
C 27	68.8	5.0	863	13	BI766821
C 28	67.8	4.9	635	10	BB663706
C 29	66.2	4.8	477	13	BI133726
C 30	66.2	4.8	541	14	BM933827
C 31	66.2	4.8	541	14	BM942392
C 32	66.2	4.8	806	13	BI916262
C 33	57.2	4.1	925	17	CNS0091P
C 34	57.2	4.1	925	17	CNS0091P
C 35	56	4.1	523	10	AW077096
C 36	55.4	4.0	925	17	CNS0091P
C 37	54.2	3.9	1101	17	CNS0039G
C 38	52.4	3.8	316	10	AW797167
C 39	52.4	3.8	1010	14	BQ668286
C 40	51.8	3.7	1125	9	AL547503
C 41	51.4	3.7	642	10	AV609566
C 42	51.2	3.7	319	9	AU257993
C 43	51.2	3.7	843	17	CNS01X1K
C 44	51	3.7	957	17	CNS015W7
C 45	50.8	3.7	354	14	BQ298642

ALIGNMENTS

RESULT 1  
BQ953400  
LOCUS  
DEFINITION  
AGENCOURT\_8784199 NIH\_MGC\_43 Homo sapiens cDNA clone IMAGE:6376296  
5', mRNA sequence.  
ACCESSION  
BQ953400  
VERSION  
BQ953400.1  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 879)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
AUTHORS  
TITLE  
National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL  
Unpublished (1999)  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabs-r@mail.nih.gov](mailto:cgabs-r@mail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
Plate: LLCM2558 row: e column: 01  
High quality sequence stop: 690.  
FEATURES  
source  
1. .879

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6376296"  
/clone\_lib="NIH\_MGC\_43"  
/tissue\_type="normal pigmented retinal epithelium"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: eye; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGCAGGAG(G). Library constructed by Ling Hong  
in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library. |"  
BASE COUNT 194 a 275 c 236 g 173 t 1 others  
ORIGIN

Query Match 56.2%; Score 776.2; DB 14; Length 879;  
Best Local Similarity 98.8%; Pred. No. 2.7e-137;  
Matches 792; Conservative 0; Mismatches 9; Indels 1; Gaps 1;  
QY 524 AAATGGGTGGGGTGGGGAGACGACTCACTTGGATTTCTTGAAGGCTTCCTGGCCCTC 583  
DB 1 AAATGGGTGGGGTGGGGAGACGACTCACTTGGATTTCTTGAAGGCTTCCTGGCCCTC 60  
QY 584 CAGCCACGTAATTCGCCCGCTCTGGATCTGGTCTAGCTTCGGGATTCGGTGGCCAGTCC 643  
DB 61 CAGCCACGTAATTCGCCCGCTCTGGATCTGGTCTAGCTTCGGGATTCGGTGGCCAGTCC 120  
QY 644 GCGGGGTAGATGTTCTGACGGGCCCCAAGGGTGCCTGAACGGCGCGGTCACTCTCT 703  
DB 121 GCGGGGTAGATGTTCTGACGGGCCCCAAGGGTGCCTGAACGGCGCGGTCACTCTCT 180  
QY 704 TCAGGAAGACTTCGAAGTGGGACACTTCTTCTCATGGATGACGACGGCGCCCGCGT 763  
DB 181 TCAGGAAGACTTCGAAGTGGGACACTTCTTCTCATGGATGACGACGGCGCCCGCGT 240  
QY 764 AGAAGGGTCCCGTGGGGTACAAAGACGCTCTTCCAGCAGGGGTGAGACAGGTGGC 823  
DB 241 AGAAGGGTCCCGTGGGGTACAAAGACGCTCTTCCAGCAGGGGTGAGACAGGTGGC 299  
QY 824 TGGACTGGCGTGTGCGGCTATCTTCCCGCTGGCGCGCGCTCAGCTGCTGCTTC 883  
DB 300 TGGACTGGCGTGTGCGGCTATCTTCCCGCTGGCGCGCGCTCAGCTGCTGCTTC 359  
QY 884 GGTGGGAGGACCTTCGCTCTCCAGGGCTCACCAGCCAGGCGCGGGATCGCC 943  
DB 360 GGTGGGAGGACCTTCGCTCTCCAGGGCTCACCAGCCAGGCGCGGGATCGCC 419  
QY 944 TCCTGAACGAAACGAAACTGACGAATCCACAGGTGAAGAGAAGTAACGGCCGCTGGC 1003  
DB 420 TCCTGAACGAAACGAAACTGACGAATCCACAGGTGAAGAGAAGTAACGGCCGCTGGC 479  
QY 1004 CTAGGGCTCCACCGAGAGAGACTAGGAGCTTGCAGAGCTCGGAGTAGAGCTCAAGT 1063  
DB 480 CTAGGGCTCCACCGAGAGAGACTAGGAGCTTGCAGAGCTCGGAGTAGAGCTCAAGT 539  
QY 1064 TTTTACCGTGGCTGCACAGCAATCAGGACCGGCTGCGGACACACACAGGTTCA 1123  
DB 540 TTTTACCGTGGCTGCACAGCAATCAGGACCGGCTGCGGACACACACAGGTTCA 599  
QY 1124 CTGTCTACGGGAGAAATCAAGGTGGACAGCTTCTGAGCAGGAGCGGAAACGCGGGGC 1183  
DB 600 CTGTCTACGGGAGAAATCAAGGTGGACAGCTTCTGAGCAGGAGCGGAAACGCGGGGC 659  
QY 1184 CTTCAACAGGACGCGCTAGTAGGGCAGAGAGAGGAGGACGACACACACACAC 1243  
DB 660 CTTCAACAGGACGCGCTAGTAGGGCAGAGAGAGGAGGAGGACGACACACACAC 719  
QY 1244 ACAAAATATGGTGAACCCCAATTTCTTACATCATATCTGTGCTACCCCTTTCCAAACAGCCT 1303  
DB 720 ACAAAATATGGTGAACCCCAATTTCTTACATCATATCTGTGCTACCCCTTTCCAAACAGCCT 779

QY 1304 AAATTTCTTTTCTCTCTCTCTT 1325  
DB 780 TAATTTCTCTTCTCTCTCTT 801  
RESULT 2  
BG749180 1045 bp mRNA linear EST 15-MAY-2001  
LOCUS 602708139F1 NIH\_MGC\_43 Homo sapiens cDNA clone IMAGE:4844744 5',  
DEFINITION mRNA sequence.  
ACCESSION BG749180  
VERSION BG749180.1 GI:14059833  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1045)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-re@mail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LCM1682 row: n column: 09  
High quality sequence stop: 822.  
FEATURES  
source  
1..1045  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4844744"  
/clone\_lib="NIH\_MGC\_43"  
/tissue\_type="normal pigmented retinal epithelium"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: eye; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGCAGGAG(G). Library constructed by Ling Hong  
in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library. |"  
BASE COUNT 245 a 322 c 300 g 178 t  
ORIGIN  
Query Match 52.8%; Score 730.2; DB 12; Length 1045;  
Best Local Similarity 95.5%; Pred. No. 1.3e-128;  
Matches 828; Conservative 0; Mismatches 28; Indels 11; Gaps 7;  
QY 522 CTAATGGTGGGGTGGGGAGACCGGACTCACTTGGATTTCTTGAAGGCTTCCTGGCC 581  
DB 2 CTAATGGTGGGGTGGGGAGACCGGACTCACTTGGATTTCTTGAAGGCTTCCTGGCC 61  
QY 582 TCCAGCCAGGTAATTCGCCCGCTCTGGATCTGGTCTAGCTTCCGGATTCGGTGGCCAGT 641  
DB 62 TCCAGCCAGGTAATTCGCCCGCTCTGGATCTGGTCTAGCTTCCGGATTCGGTGGCCAGT 121  
QY 642 CCGCGGGGTAGATGTTCTTCCAGCGGCCCCAAAGGGTGCCTGAACGCCCGCGGTACCTTC 701  
DB 122 CCGCGGGGTAGATGTTCTTCCAGCGGCCCCAAAGGGTGCCTGAACGCCCGCGGTACCTTC 181  
QY 702 CTTCAGGAAGACTTCGAAGCTGGACACCTTCTTCTCATGGATGACACGCGGGCGCCCCG 761  
DB 182 CTTCAGGAAGACTTCGAAGCTGGACACCTTCTTCTCATGGATGACACGCGGGCGCCCCG 241  
QY 762 GTAGAAGGGGTCCCGTTCGGGTACACAGCAGGCTCTTCACGACGCGGCTGACACAGGTC 821  
|||||

Db	242	GTAGAAGGGTCCCGCTTCGGGTACAAAGCACGCTCTTCAGCAC - GCCTGAGACAGGTG	300
Qy	822	GCTGGACCTGGCGTCTGCTCCGCTCATCTTCCCGCTGGCGCGGCTCAGCTCGCTGCT	881
Db	301	GCTGGACCTGGCGTCTGCTCCGCTCATCTTCCCGCTGGCGCGGCTCAGCTCGCTGCT	360
Qy	882	TCCGCTGGGAGGCACCTCCGCTGTCCAGCGGCTCACGGACCCAGGGCGCGGATCG	941
Db	361	TCGCGTGGGAGGCACCTCCGCTGTCCAGCGGCTCACCGACCCA - GGC GCGGGATCG	419
Qy	942	CCTCTGAAACGAACGAGAAATGACGAATCCACAGGTGAAGAGAGCTAACGCCGCTGC	1001
Db	420	CCTCTGAAACGAACGAGAAATGACGAATCCACAGGTGAAGAGAGATACGCCGCTGC	479
Qy	1002	GCCTAGGCGCTCACCCAGAGGAGACACTAGGAGCTTCGAGGACTCGGAGTAGACGCTCAA	1061
Db	480	GCCTAGGCGCTCACCCAGAGGAGACACTAGGAGCTTCGAGGACTCGGAGTAGACGCTCAA	539
Qy	1062	GTTTTTCACGTTGGGTGCA - CAGCAGTACAGGACCCGAGTGGCGGCACACACAGGT	1120
Db	540	GTTTTTCACCATGGCTGCAACGAGCCAATCAGGAGCCGACGTGGCGGCACACACAGGT	599
Qy	1121	TCACCTGCTACGGGCAGAATCAAGGTGGACACTTCTGAGCAGGAGCCGGAAACGCGCGG	1180
Db	600	TCACCTGCTACGGGCAGAAATCAAGGTGGACACTTCTGAGCAGGAGCCGGAAACGCGCGG	659
Qy	1181	GGCCTTCAACAGGCACGCTTAGTGAGGCGAGGAGGAGGAGCACACA - - - - CACA	1236
Db	660	GG - CTTCAACAGGCACGCTTAGTGAGGCGAGGAGGAGGAGGAGCACACAGAA	718
Qy	1237	CACACACACAAATATGGTGAACCCAAATTCATTACATCATATCTGTGTGTACCCCTTCCAA	1296
Db	719	CACACACAAATATGGTGAACCCAAATTCATTACATCATATCTGTGTGTACCCCTTCCAA	778
Qy	1297	ACAGCCTAAATTTTCTTTTCTCTCTTCTGTGCACTTTTACCCCTC - AATCTCCTGCTTCCT	1355
Db	779	ACAGC - - TAATTTTCTTTTCTCTCTTCTGTGCACTTTTACCCCTCAAAATCTCCTGATGCT	836
Qy	1356	CCCAAAATTAAGCAATTAAGTTCTCTGG	1382
Db	837	CCCAAAATTAAGCAATTAAGTTCTCTGG	863

**RESULT 3**  
**LOCUS** BQ948660/c  
**DEFINITION** ACENGCURT\_8784237 NIH\_MGC\_43 Homo sapiens cDNA clone IMAGE:6376418  
**ACCESSION** BQ948660  
**VERSION** EST.  
**KEYWORDS** 5', mRNA sequence.  
**SOURCE** BQ948660.1 GI:22364138  
**ORGANISM** human.  
**REFERENCE** Homo sapiens  
**AUTHORS** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
**TITLE** Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
**JOURNAL** NIH-MGC <http://mgc.nci.nih.gov/>.  
**COMMENT** 1 (bases 1 to 845)  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabps-remail.nih.gov](mailto:cgabps-remail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: L1CM2558 row: j column: 03  
High quality sequence stop: 534.  
Location/Qualifiers  
1..845  
/organism="Homo sapiens"

```

/db_xref="taxon:9606"
/clone="IMAGE:6376418"
/clone_lib="NIH_MGC_43"
/tissue_type="normal pigmented retinal epithelium"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pORF7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: This is a NIH_MGC Library. |"
184 a 219 c 255 g 157 t 30 others
BASE COUNT
ORIGIN

```

	FEATURES	SOURCE
1.	100% Pure Cotton	USA
2.	Machine Washable	India
3.	Soft & Comfortable	Bangladesh
4.	Durable & Long Lasting	Pakistan
5.	Eco-Friendly	China
6.	Available in Multiple Colors	Vietnam
7.	Lightweight & Breathable	Thailand
8.	Easy to Fold & Store	Philippines
9.	Hypoallergenic	Malaysia
10.	Perfect for Travel	Singapore

```

RESULT 4
BQ650550/c
LOCUS
DEFINITION BQ650550 898 bp mRNA linear EST 15-JUL-2002
AGENCOURT_8350342 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6286493
5', mRNA sequence.
ACCESSION BQ650550
VERSION BQ650550.1 GI:21774722
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: CGAP (Stanford)
cDNA Library Preparation: Ruben Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2487 row: g column: 06
High quality sequence stop: 667.
Location/Qualifiers
1..898
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6286493"
/clone_lib="NIH_MGC_100"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GCGACGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."
BASE COUNT 206 a 252 c 273 g 167 t
ORIGIN
Query Match. 44.1%; Score 609.6; DB 14; Length 898;
Best Local Similarity 98.2%; Pred. No. 9.8e-106;
Matches 648; Conservative 0; Mismatches 9; Indels 3; Gaps 3;
QY 292 TTTTATTTCTTAAGATTCTCTGTT-GTTTATACAGATTTTAA-GTTTACTCTTACT 349
DB 771 TTTTATTTCTTAAGGATTCCCGTTGGTTTATACAGATTTTAAAGTTTACTCTTACG 712
QY 350 GCTGACCAAGTGAAATCTCTCCAGTCACAGTGCAACCTCTA-CCCCCAACTGCA 408
DB 711 GCTGACCAAGTGAAATCTCTCTCCAGTCACAGGTCACCTTACCCTCCCAACTGCA 652
QY 409 ACAGAGTTTGTAGGGGCATCAATCACACCGAGAAGTCACAGCCCTCAACCACTGAGGT 468
DB 651 ACAGAGTTTGTAGGGGCATCAATCACACCGAGAAGTCACAGCCCTCAACCACTGAGGT 592
QY 469 GTGGGGGGTAGGATCTGATTTCTTCATATCAACCCACACATATAGGCGACCTAAATG 528
DB 591 GTGGGGGGTAGGATCTGATTTCTTCATATCAACCCACACATATAGGCGACCTAAATG 532
QY 529 GTGGGGGGTAGGAGACCGACTCACTTCAGTTTCTTGAAGGCTTCTGGCTCCAGGC 588
DB 531 GTGGGGGGTAGGAGACCGACTCACTTCAGTTTCTTGAAGGCTTCTGGCTCCAGGC 472
QY 589 ACCTAAATGCCCCGCTCTGGATCTGGTCTAGCTTCGGGATTCGGTGGCCAGTCCCGGG 648
|||||

```

```

DB 471 ACCTAAATGCCCCGCTCTGGATCTGGTCTAGCTTCGGATTCGGTGGCCAGCTCCGCGG 412
QY 649 GTCTAGATGTTCTCTGAGGGCCCCAAAGGGTGCCTGAACGCGCGGTCACTCTCTTCAGG 708
|||||
DB 411 GTCTAGATGTTCTCTGAGGGCCCCAAAGGGTGCCTGAACGCGCGGTCACTCTCTTCAGG 352
QY 709 AAGACTTCGAAGCTGGACACCTTCTTCTCATGATGACGACGCGGCGGCCCGGTAGGAAG 768
|||||
DB 351 AAGACTTCGAAGCTGGACACCTTCTTCTCATGATGACGACGCGGCGGCCCGGTAGGAAG 292
QY 769 GGTTCCTCCGTTGGGTATACAAAGACGCTCTTACAGAGGGCTTGAACAGGTGGCTGGAC 828
|||||
DB 291 GGTTCCTCCGTTGGGTATACAAAGACGCTCTTACAGAGGGCTTGAACAGGTGGCTGGAC 232
QY 829 CTGCGCTGCTGCGGCTCATCTTCCCGGCTGGCGGCTCAGCTCGCTTCGGGCTC 888
|||||
DB 231 CTGCGCTGCTGCGGCTCATCTTCCCGGCTGGCGGCTCAGCTCGCTTCGGGCTC 172
QY 889 GGGAGGCACCTCCGCTGTCCAGCGGCTCACCGACCCAGCGGCGGGGATCGCTCTCTG 948
|||||
DB 171 GGGAGGCACCTCCGCTGTCCAGCGGCTCACCGACCCAGCGGCGGGGATCGCTCTCTG 112
RESULT 5
BQ650598/c
LOCUS
DEFINITION BQ650598 963 bp mRNA linear EST 15-JUL-2002
AGENCOURT_8207577 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6283187
5', mRNA sequence.
ACCESSION BQ650598
VERSION BQ650598.1 GI:21774770
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: CGAP (Stanford)
cDNA Library Preparation: Ruben Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2478 row: m column: 12
High quality sequence stop: 618.
Location/Qualifiers
1..963
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6283187"
/clone_lib="NIH_MGC_100"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GCGACGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."
BASE COUNT 171 a 350 c 244 g 197 t 1 others
ORIGIN
Query Match 38.7%; Score 534.4; DB 14; Length 963;
Best Local Similarity 97.1%; Pred. No. 1.7e-91;
Matches 544; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

```



[illegible]

QY	824	TGGACCTGGCGCTGCTGCCGCTCATCTTTCCTCCGCTGCCGCGCCCTCAGCTCGCTCGCTTC	883
Db	225	TGGACCTGGCGCTGCTGCCGCTCATCTTTCCTCCGCTGCCGCGCCCTCAGCTCGCTCGCTTC	
QY	884	GCCTCGGAGGACCACTCCGGTGTCCCGAGCGGCTCACCAGCACCCAGGCGCGGATCGCC	943
Db	165	GCCTCGGAGGACCACTCCGGTGTCCCGAGCGGCTCACCAGCACCCAGGCGCGGATCGCC	106
QY	944	TCCTG 948	
Db	105	TCCTG 101	
RESULT 8			
BM923267/c			
LOCUS	BM923267	984 bp	mRNA linear EST 12-MAR-93
DEFINITION	AGENCOURT_6626059 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:57557		
ACCESSION	BM923267		
VERSION	BM923267.1		
KEYWORDS	EST		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	1 (bases 1 to 984)		
TITLE	NIH-MGC http://mgc.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: Life Technologies, Inc. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: LLAM12803 row: e column: 22 High quality sequence stop: 673.		
FEATURES	Location/Qualifiers		
source	1..984		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:5758845"		
	/clone_lib="NIH_MGC_116"		
	/lab_host="DH10B"		
	/note="Organ: pooled colon, kidney, stomach; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed) source anonymous pool of 3 colons, age 26 yo male, 4 female, 71 yo male colon; 46 yo male kidney, and pool stomachs, 62 yo male and 70 yo female. Library is oligo-dT primed and directionally cloned (EcoRV site destroyed upon cloning). Average insert size 1.4 kb. insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by Gruber (Invitrogen). Research Genetics tracking code 023. Note: this is a NIH_MGC Library."		
BASE COUNT	246 a	254 c	279 g
ORIGIN	205 t		
Query Match	28.4%; Score 391.8; DB 14; Length 984;		
Best Local Similarity	99.5%; Pred. No. 1.7e-64;		
Matches 393;	Conservative	0;	Mismatches 2; Indels 0; Gaps
QY	554	ACTTGAGTTTCTGAAGGCTTCCTGGCGCTCCAGCCACGTAATTCGCCCGCTCTGGATCT	61
Db	517	AATTGAGTTTCTTGAAGGCTTCCTGGCGCTCCAGCCACGTAATTCGCCCGCTCTGGATCT	45
QY	614	GGTCTAGCTCCGGATTCCGTTGGCGCAGTCGCGGGGTGTAGATGTTTCCTGACGGCCCCAA	67
Db	457	GGTCTAGCTCCGGATTCCGTTGGCGCAGTCGCGGGGTGTAGATGTTTCCTGACGGCCCCAA	39



Qy	674	AGGTGGCTGAACGCCCGCGGTCAACCTCCCTTCAGGAAGACTTCGAAGCTGGACACCTTCT	733
Db	397	AGGTGGCTGAACGCCCGCGGTCAACCTCCCTTCAGGAAGACTTCGAAGCTGGACACCTTCT	338
Qy	734	TCTCATGGATGACACGCGCGGCCCGGTAGAGGGGTCCCGTTCGGGTACACAAGCA	793
Db	337	TCTCATGGATGACACGCGCGGCCCGGTAGAGGGGTCCCGTTCGGGTACACAAGCA	278
Qy	794	CGCTTTTCAGACGCGGTGAGACAGGTGGCTGGACCTGGCGCTGCTCGCGCTCATCTTCC	853
Db	277	CGCTTTTCAGACGCGGTGAGACAGGTGGCTGGACCTGGCGCTGCTCGCGCTCATCTTCC	218
Qy	854	CCGTGGCGCGCGCTCAGCTCGCTGCTTCGGCTCGGGAGGCACCTCGCTGTCCACGC	913
Db	217	CCGTGGCGCGCGCTCAGCTCGCTGCTTCGGCTCGGGAGGCACCTCGCTGTCCACGC	158
Qy	914	GCCTCAGCGCACCCAGCGCGGGATCGCTTCCTG	948
Db	157	GCCTCAGCGCACCCAGCGCGGGATCGCTTCCTG	123
RESULT 9			
LOCUS	BQ644671	929 bp	linear
DEFINITION	AGENCOURT_8490723 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6293873		EST 15-JUL-2002
ACCESSION	BQ644671		
VERSION	BQ644671.1	GI:21768843	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 929)		
TITLE	NIH-MGC http://mgc.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgabbs-remail.nih.gov		
	Tissue Procurement: CGAP (Stanford)		
	cDNA Library Preparation: Rubin Laboratory		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)		
	DNA Sequencing by: Agencourt Bioscience Corporation		
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov		
	Plate: LLCM2497 row: j column: 18		
	High quality sequence stop: 582.		
FEATURES	Location/Qualifiers		
source	1..929		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:6293873"		
	/clone_lib="NIH_MGC_100"		
	/tissue_type="hepatocellular carcinoma, cell line"		
	/lab_host="DH10B (phage-resistant)"		
	/note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(C). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."		
BASE COUNT	283 a	238 g	196 t
ORIGIN			
Query Match	20.9%	Score 288.4;	DB 14; Length 929;
Best Local Similarity	99.7%	Pred. No. 6.7e-45;	
Matches 289; Conservative	0;	Mismatches 1;	Indels 0; Gaps 0;

Db	290	AATTGAGTTTCTTGAAGGCTTCTGGGCTTCAGCCAGTAATTCGCCCGCTCTGGATCT	231
QY	614	GGTCTAGCTTCGGGATTTCGGTGGCCAGTCCGCGGGGTGTAGATTCTCTGACGGCCCAA	673
Db	230	GGTCTAGCTTCGGGATTTCGGTGGCCAGTCCGCGGGGTGTAGATTCTCTGACGGCCCAA	171
QY	674	AGGTGCTCTGAAGCCGCCCGTCCACCTCTTCAGGAAGACTTCGAAGCTGACACCTTCT	733
Db	170	AGGTGCTCTGAAGCCGCCCGTCCACCTCTTCAGGAAGACTTCGAAGCTGACACCTTCT	111
QY	734	TCTCATGATGACACGCGGGGCCCGCTAGAGGGGTCCCGTTCGGGTACACAAGCA	793
Db	110	TCTCATGATGACACGCGGGGCCCGCTAGAGGGGTCCCGTTCGGGTACACAAGCA	51
QY	794	CGCTTTCACGACGGGCTGAGACAGGTGGCTGGACCTGGCGCTGCTGCCG	843
Db	50	CGCTTTCACGACGGGCTGAGACAGGTGGCTGGACCTGGCGCTGCTGCCG	1
RESULT 11			
BM927835/c			
LOCUS	BM927835	1019 bp	mrna linear EST 12-MAR-2002
DEFINITION	AGENCOURT_6729788 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:5797671		
	5', mRNA sequence.		
ACCESSION	BM927835		
VERSION	BM927835.1	GI:19378214	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 1019)		
TITLE	NTH-MGC http://mgc.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgapbs-remail.nih.gov		
	Tissue Procurement: CGAP (Stanford)		
	cDNA Library Preparation: Rubin Laboratory		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)		
	DNA Sequencing by: Agencourt Bioscience Corporation		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LLNL at:		
	http://image.llnl.gov		
	Plate: LNCM2024 row: 9 column: 16		
	High quality sequence stop: 750.		
FEATURES	Location/Qualifiers		
source	1..1019		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:5797671"		
	/clone_lib="NIH_MGC_100"		
	/tissue_type="hepatocellular carcinoma, cell line"		
	/lab_host="DH10B (phage-resistant)"		
	/notes="Organ: liver; Vector: pOT87; Site:1: XhoI; Site:2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(5). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."		
BASE COUNT	315 a 229 c 258 g 217 t		
ORIGIN			
Query Match	20.9%	Score 288.4; DB 14; Length 1019;	
Best Local Similarity	99.7%	Pred. No. 6.6e-45;	
Matches	289; Conservative	0; Mismatches 1; Indels 0; Gaps 0;	
QY	554	ACTTGAGTTTCTTGAAGGCTTCTGGGCTCCAGCCAGTAATTCGCCCGCTCTGGATCT	613



REFERENCE  
AUTHORS  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 364)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR2&t2=MR2-NN111-  
070800-007-hl2&t3=2000-08-07&t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 9.  
Location/Qualifiers  
1. .364  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="NN111"  
/dev\_stage="Adult"  
/note="Organ: nervous\_normal; Vector: puc18; Site:1: SmaI;  
Site:2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the puc 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."  
82 a 114 c 112 g 56 t

## FEATURES

source  
1. .364  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="NN111"  
/dev\_stage="Adult"  
/note="Organ: nervous\_normal; Vector: puc18; Site:1: SmaI;  
Site:2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the puc 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."  
82 a 114 c 112 g 56 t

## BASE COUNT

ORIGIN

Query Match 18.1%; Score 250.8; DB 12; Length 364;  
Best Local Similarity 94.3%; Pred. No. 1.2e-37;  
Matches 315; Conservative 0; Mismatches 12; Indels 7; Gaps 5;

QY 860 GCGGCGGCTCAGCTCGCTTGGC-CTCGGAGGACCTCCGCTGTCACGAGGCGCTC 918  
Db 15 GGCTCGGCTCAGCTCGCTTGGCTGCTGGGAGGACCTCCGCTGTCACGAGGCGCTC 74  
QY 919 ACCGACCCAG-GGGCGGGATCGCTCTGAAAGCAAGCAACTGACGAATCCACAG 977  
Db 75 ACCGACCCAGTGGCGGGATCGCTCTGAAAGCAAGCAACTGACGAATCCACAG 134  
QY 978 GTGAAGAAGAAGTAACGGCCGCTGCGCTAGCGCTACCCAGAGGAGACACTAGGAGCTT 1037  
Db 135 GTGAAGAAGAAGTAACGGCCGCTGCGCTAGCGCTACCCAGAGGAGACACTAGGAGCTT 194  
QY 1038 GCAGACTCGGAGTAGAGCTCAAGTTTTCACCGTGGCGTGCA--CAGCCAAATCAGG-- 1093  
Db 195 GCAGACTCGGAGTAGAGCTCAAGTTTTCACCGTGGCGTGCAACATCCAGGGA 254  
QY 1094 ACCCGACTGGCGGACACACACAGGTTTCACCTAGCGGAGCAATCAAGTGGAC-AG 1152  
Db 255 CCGGAGTTGGCGGACACACACAGGTTTCACCTAGCGGAGCAATCAAGTGGACAG 314  
QY 1153 TTCTGTAGCAGGAGCGGAAACGCGGGGCGCTT 1186  
Db 315 TTCTGTAGCAGGAGCGGAAACGCGGGGCGCTT 348

RESULT 15  
AA863443

## LOCUS

DEFINITION  
AA863443  
chr05e10.s1 NCI\_CGAP\_kid3 Homo sapiens cdna clone IMAGE:1456938 3',  
mRNA sequence.  
AA863443  
VERSION  
AA863443.1 GI:2955922  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens

## REFERENCE

1 (bases 1 to 246)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Unpublished (1997)  
Tumor Gene Index

## AUTHORS

CONTACT: Robert Strausberg, Ph.D.  
Email: cgapbs-f@mail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.

## JOURNAL

CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

## COMMENT

Insert length: 1358 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 208.  
Location/Qualifiers

## FEATURES

source  
1. .246  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1456938"  
/clone\_lib="NCI\_CGAP\_kid3"  
/lab\_host="DH10B"  
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with  
a modified polylinker; Site:1: Not I; Site:2: Eco RI; 1st  
strand cDNA was primed with a Not I - Oligo(dT) primer,  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not  
I and Eco RI sites of the modified pT73 vector. mRNA  
source: 2 pooled kidneys. Library went through one round  
of normalization. Library constructed by Bento Soares and  
M. Fatima Bonaldo."  
63 a 66 c 44 g 73 t

## BASE COUNT

ORIGIN

Query Match 17.7%; Score 244.4; DB 9; Length 246;  
Best Local Similarity 99.6%; Pred. No. 2.2e-36;  
Matches 245; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 268 GTCACCTACCTTTCGAAGGATACCTTTTATTTTCTTTAAGATTCTCTGTTTATACAC 327  
Db 1 GTCATTACCTTTCGAAGGATACCTTTTATTTTCTTTAAGATTCTCTGTTTATACAC 60  
QY 328 AGATTTTAAAGTTTACTCTCTACTGTCACCCAAAGTGAATTCCTTCACAGTCACAGTGC 387  
Db 61 AGATTTTAAAGTTTACTCTCTACTGTCACCCAAAGTGAATTCCTTCACAGTCACAGTGC 120  
QY 388 AACCTCTACCCCCCACTGCAACGAGAGTTTGGGGGCATCAATCACACCGAGAAGTCA 447  
Db 121 AACCTCTACCCCCCACTGCAACGAGAGTTTGGGGGCATCAATCACACCGAGAAGTCA 180  
QY 448 CAGCCCTCAACCACTGAGGTGGGGGGTAGGGATCTGCATTTCATATCAACCCC 507  
Db 181 CAGCCCTCAACCACTGAGGTGGGGGGTAGGGATCTGCATTTCATATCAACCCC 240  
QY 508 ACACATA 513  
Db 241 ACACATA 246

Search completed: November 26, 2002, 11:58:01

Job time : 2077 secs

---

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 25, 2002, 08:44:00 ; Search time 265 Seconds  
(without alignments)  
11744.384 Million cell updates/sec

Title: US-09-674-593-1

Perfect score: 1382

Sequence: 1 cattatgctaacagcataaa.....taaagcaattaagtctctg 1382

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_101002.\*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*
- 24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1382	100.0	1382	21	AAZ36643 Human tumour rejec
C 2	595.4	43.1	2167	21	AAZ36644 Human tumour rejec
C 3	550	39.8	1956	22	AAI58192 Human polynucleoti
C 4	420.6	30.4	643	24	ABQ42610 Oligonucleotide fo
5	420.6	30.4	643	24	ABQ42611 Oligonucleotide fo
6	398.8	28.9	643	24	ABQ42612 Oligonucleotide fo
C 7	398.8	28.9	643	24	ABQ42613 Oligonucleotide fo
8	244.4	17.7	246	21	AAZ36649 EST AA863443 which
C 9	139.4	10.1	492	22	AAI59978 Human polynucleoti

C 10	68.4	4.9	5284	23	AAZ36643	DNA encoding novel
C 11	68	4.9	2424	22	AAZ36643	Novel protein kina
C 12	60	4.3	60	24	ABN38745	Human spliced tran
C 13	51.6	3.7	1645	21	AAZ36643	Human breast and o
C 14	51.6	3.7	11735	22	AAZ36643	Chemically pretrea
C 15	51.6	3.7	11735	24	AAZ36643	Human gene regulat
C 16	51.6	3.7	11735	24	ABK28167	DNA transcription
C 17	49.4	3.6	1937	23	ABV23182	Human prostate exp
C 18	48.4	3.5	823	22	ABA07664	Human ovarian and
C 19	48.4	3.5	823	22	AAZ36643	Human reproductive
C 20	48.4	3.5	3509	21	AAZ36643	Arabidopsis RANL C
C 21	48.4	3.5	4800	21	AAZ36643	Arabidopsis RANL g
C 22	48	3.5	14987	24	ABL32630	Human immune syste
C 23	47.2	3.4	2482	24	ABK35668	cDNA sequence #59
C 24	47.2	3.4	513445	22	AAI61373	Soybean 318013 reg
C 25	46.4	3.4	353	22	AAI61373	Human breast cance
C 26	46	3.3	6092	24	AAZ36643	Human gene regulat
C 27	45.8	3.3	1754	13	AAQ20239	Gene encoding cell
C 28	45.4	3.3	5908	22	AAZ36643	Chemically pretrea
C 29	45.4	3.3	5908	24	AAZ36643	Human gene regulat
C 30	45.4	3.3	5908	24	ABK28231	DNA transcription
C 31	45	3.3	2991	22	AAI5458	Human scavenger re
C 32	45	3.3	3019	22	AAI5458	Human scavenger re
C 33	45	3.3	3900	24	ABK15178	Human REPTR 10 CDN
C 34	44.8	3.2	1212	20	AAZ36643	Human WISP-3 prote
C 35	44.8	3.2	1212	20	AAZ36643	Human WISP-3 prote
C 36	44.6	3.2	1967	21	AAZ36643	Human cancer assoc
C 37	44.6	3.2	3265	22	AAZ36643	Human secreted pro
C 38	44.6	3.2	3265	22	AAZ36643	Human secreted pro
C 39	44.6	3.2	3339	22	AAZ36643	Human secreted pro
C 40	44.6	3.2	3339	22	AAZ36643	Human secreted pro
C 41	44.4	3.2	1159	22	AAZ36643	Human secreted pro
C 42	44.2	3.2	245	23	ABV07448	Human prostate exp
C 43	44.2	3.2	6106	22	AAZ36643	Tumour suppressor
C 44	44.2	3.2	6106	24	ABK40031	Human chemically p
C 45	44.2	3.2	6106	24	ABL33472	Human immune syste

ALIGNMENTS

RESULT 1  
AAZ36643  
ID AAZ36643 standard; cDNA; 1382 BP.

XX AC AAZ36643;

XX DT 22-FEB-2000 (first entry)

XX DE Human tumour rejection antigen RUR-1 antisense cDNA sequence.

XX KW Antisense; human; tumour rejection antigen; RUR-1; tumour; cancer;

XX KW renal cell carcinoma; colorectal carcinoma; melanoma; sarcoma;

XX KW leukaemia; ss.

XX OS Homo sapiens.

XX XX

XX Key Location/Qualifiers

XX CDS 738..992

XX FT /\*tag= a

XX FT primer\_bind 523..547

XX FT /\*tag= b

XX FT /note= "binding site for primer VDE119 (see AAZ36647)"

XX FT complement (1280..1305)

XX FT /\*tag= c

XX FT /note= "binding site for primer VDE120 (see AAZ36648)"

XX XX WO9958546-A1.

XX XX 18-NOV-1999.

XX XX

XX PF 13-MAY-1999; 99WO-US10424.

XX XX

PR 13-MAY-1998; 98US-0085318.  
XX (LUDW-) LUDWIG INST CANCER RES.  
PA  
XX  
XX  
PI Van Den Eynde B, Boon-Falleur T;  
XX  
XX WPI; 2000-053076/04.  
DR P-PSDB; AAY53809.  
XX  
XX New isolated tumour rejection antigen RUR-1 nucleic acids, used for,  
PT e.g. treatment of cancers -  
XX  
XX Claim 4; Fig 5; 75pp; English.  
XX  
XX The present sequence represents the antisense cDNA sequence of human  
CC tumour rejection antigen RUR-1. The present sequence is the antisense  
CC strand of a ubiquitously expressed gene. The antisense strand codes for  
CC a polypeptide which is preferentially expressed in tumour samples and  
CC tumour-derived cells lines. The polypeptide is unrelated to any TRAP  
CC protein. The sequence was isolated from a renal cell carcinoma line  
CC LB9211-RC. The RUR-1 nucleic acids and polypeptides can be used for  
CC diagnosis, prognosis or treatment of a disorder characterized by the  
CC expression of a RUR-1 antisense cDNA molecule or an expression product,  
CC such as cancers, e.g. renal cell carcinoma, colorectal carcinoma,  
CC melanoma, sarcoma or leukaemia.  
XX  
SQ Sequence 1382 BP; 355 A; 373 C; 344 G; 310 T; 0 other;

Query Match 100.0%; Score 1382; DB 21; Length 1382;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1382; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CATATGCTAACAGCATAAACATGAGGGGTGGAGCAGGGTCCAAAAAGTGGTGTG 60  
DB 1 CATATGCTAACAGCATAAACATGAGGGGTGGAGCAGGGTCCAAAAAGTGGTGTG 60  
QY 61 TCAATTCCTGGAATGAAGTGAATTAATTAACAGTACGGGAATGCAGAGCAA 120  
DB 61 TCAATTCCTGGAATGAAGTGAATTAATTAACAGTACGGGAATGCAGAGCAA 120  
QY 121 TTTTCTCTCTGTCACAAATAGTGTCCACACTTGAAGTGAATTAAGAAATGTTA 180  
DB 121 TTTTCTCTCTGTCACAAATAGTGTCCACACTTGAAGTGAATTAAGAAATGTTA 180  
QY 181 TTTAAATTAAGAGTGGATTTCCAGGAAAAAATAAGAAAAAGGAAAAAAACTG 240  
DB 181 TTTAAATTAAGAGTGGATTTCCAGGAAAAAATAAGAAAAAGGAAAAAAACTG 240  
QY 241 AACAGAAACGCAAAAGTATCAGTTTGGTCACTAACCTTTGCAAGGATACCTTTTATT 300  
DB 241 AACAGAAACGCAAAAGTATCAGTTTGGTCACTAACCTTTGCAAGGATACCTTTTATT 300  
QY 301 TCTTTAAGATTCCTGTGTTTATACAGATTTTAAGTTTACTCTACTGTGACCCAAAG 360  
DB 301 TCTTTAAGATTCCTGTGTTTATACAGATTTTAAGTTTACTCTACTGTGACCCAAAG 360  
QY 361 TGAATTCCTTCCAGTCACAGTGTCAACCTCTACCCCCCACTGCAACGAGAGTTTG 420  
DB 361 TGAATTCCTTCCAGTCACAGTGTCAACCTCTACCCCCCACTGCAACGAGAGTTTG 420  
QY 421 AGGGGATCAATACACACGAGAGTACAGCCCTCAACCTAGGTGTGGGGGGTAG 480  
DB 421 AGGGGATCAATACACACGAGAGTACAGCCCTCAACCTAGGTGTGGGGGGTAG 480  
QY 481 GGATCTGCATTTCTTCATATCAACCCACACACTATAGGGCACCTTAATGGTGGGGGTGG 540  
DB 481 GGATCTGCATTTCTTCATATCAACCCACACACTATAGGGCACCTTAATGGTGGGGGTGG 540  
QY 541 GGGAGACCGACTACATGAGTTTCTTGAAGGTTCTCGGCTCCAGCCACGTAATTGCCC 600  
DB 541 GGGAGACCGACTACATGAGTTTCTTGAAGGTTCTCGGCTCCAGCCACGTAATTGCCC 600  
QY 601 CCGCTCTGGATCTGGTCTAGCTTCGGATTCGGTGCCAGTCCGCGGGGTGTAGATGTTT 660

DB 601 CCGCTCTGGATCTGGTCTAGCTTCCGGATTCGGTGGCCAGTCCGCGGGTGTAGATGTTT 660  
QY 661 CTGACGGCCCCAAAGGGTGTCTGAACGCCCGCGGTACCTCTCTTACGAGAACTTTCGAAG 720  
DB 661 CTGACGGCCCCAAAGGGTGTCTGAACGCCCGCGGTACCTCTCTTACGAGAACTTTCGAAG 720  
QY 721 CTGGACACCTTCTTCTCATGTGATGACGACGGCGCGCGGTGAGAAAGGGTGTCCCGTTG 780  
DB 721 CTGGACACCTTCTTCTCATGTGATGACGACGGCGCGCGGTGAGAAAGGGTGTCCCGTTG 780  
QY 781 CGGTACACAAACGAGCTCTTACAGACGGGTGACAGAGGTGGCTGGACCTGGCGCTGCTG 840  
DB 781 CGGTACACAAACGAGCTCTTCTACAGACGGGTGACAGAGGTGGCTGGACCTGGCGCTGCTG 840  
QY 841 CCGCTCATCTTCCCGCTGGCGCGCTCAGCTCGCTGCTTCCGCTCGGAGGACACCTC 900  
DB 841 CCGCTCATCTTCCCGCTGGCGCGCTCAGCTCGCTGCTTCCGCTCGGAGGACACCTC 900  
QY 901 CGCTGTCCAGCGGCTCACCAGCACCCAGGCGCGGGATCGCTTCTTGAACGAGCAGAGA 960  
DB 901 CGCTGTCCAGCGGCTCACCAGCACCCAGGCGCGGGATCGCTTCTTGAACGAGCAGAGA 960  
QY 961 AACTGACGAATCCACAGGTGAAGAGAGTAACGGCGCTGGCGCTAGCGCTCCACCCAGA 1020  
DB 961 AACTGACGAATCCACAGGTGAAGAGAGTAACGGCGCTGGCGCTAGCGCTCCACCCAGA 1020  
QY 1021 GGAGACACTAGGAGCTTGCAGGACTCGGAGTAGAGCTCAAGTTTTCACCGTGGCGTGC 1080  
DB 1021 GGAGACACTAGGAGCTTGCAGGACTCGGAGTAGAGCTCAAGTTTTCACCGTGGCGTGC 1080  
QY 1081 ACAGCAATCAGGACCCGCGCTGCGGCACACACAGGTTTCACTGCTACGGGAGAAAT 1140  
DB 1081 ACAGCAATCAGGACCCGCGCTGCGGCACACACAGGTTTCACTGCTACGGGAGAAAT 1140  
QY 1141 CAAGTGACAGCTTCTGAGCAGGAGCGGAAACGCGGGGCGCTTCAACAGGACGCGCC 1200  
DB 1141 CAAGTGACAGCTTCTGAGCAGGAGCGGAAACGCGGGGCGCTTCAACAGGACGCGCC 1200  
QY 1201 TAGTGAGGCGAGGAGGAGGAGCGCACACACACACACAAATATGGTGAACCC 1260  
DB 1201 TAGTGAGGCGAGGAGGAGGAGGAGCGCACACACACACAAATATGGTGAACCC 1260  
QY 1261 CAATTTCTTACATCATATCTGTGTACCTTTTCCAAACAGCCTAAATTTCTTTCTCTC 1320  
DB 1261 CAATTTCTTACATCATATCTGTGTACCTTTTCCAAACAGCCTAAATTTCTTTCTCTC 1320  
QY 1321 TTTCTGCACCTTTTACCCCTCAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1380  
DB 1321 TTTCTGCACCTTTTACCCCTCAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1380  
QY 1381 GG 1382  
DB 1381 GG 1382

RESULT 2  
AAZ36644/c  
ID AAZ36644 standard; cDNA; 2167 BP.  
XX  
XX AAZ36644;  
XX  
XX 22-FEB-2000 (first entry)  
DT Human tumour rejection antigen RUR-1 sense cDNA sequence.  
DE  
XX Antisense; human; tumour rejection antigen; RUR-1; tumour; cancer;  
KW renal cell carcinoma; colorectal carcinoma; melanoma; sarcoma;  
KW leukaemia; ss.  
XX  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers















```
QY 676 GGTGCTGAACCGCGCGGTACCTCTCAGGAAGACTTCAAGCTGACACCTTCTTC 735
  || || || || || || || || || || || || || || || || || || || || || ||
Db 4049 GGGGCTGCACAGCTGATGCTACCTCGCAGAGGAGGCTCCTATGTTGGGAGCGGCT 3990
QY 736 TCATGATGACGACGCGCGCGCGCGGTAGAAAGGGTCCCCCTTCGGGTACAAAGCAGG 795
  || || || || || || || || || || || || || || || || || || || || || ||
Db 3989 TGAGTCACACACGCTGGGAGCGCTGGGAAGATGGTCCCCATTCGGGTACACCACTACC 3930
QY 796 CTCTTCACGACGG 809
  || || || || || || || || || || || || || || || || || || || || || ||
Db 3929 CTCTTGGCTGCTGG 3916

RESULT 11
ID AAF44647/c
AC AAF44647 standard; cDNA; 2424 BP.
XX
AC AAF44647;
XX
DT 27-MAR-2001 (first entry)
XX
DE Novel protein kinase cDNA, SEQ ID NO: 26.
XX
KW Human; mouse; protein kinase; antiarthritic; antisclerotic; osteopathic;
KW immunosuppressive; cardiac; renal; antiinflammatory; antiasthmatic;
KW dermatological; antidiabetic; antifertility; gene therapy; vaccine;
KW immune disorder; cardiovascular disease; neurodegenerative disease;
KW cancer; autoimmune disorder; stroke; inflammatory bowel disease;
KW inflammatory pelvic disease; multiple sclerosis; psoriasis; ss.
XX
OS Homo sapiens.
XX
PN WO200073469-A2.
XX
PD 07-DEC-2000.
XX
PF 26-MAY-2000; 2000WO-US14842.
XX
PR 28-MAY-1999; 99US-0136503.
XX
PA (SUGEN-) SUGEN INC.
XX
PI Plowman GD, Martinez R, Whyte D, Sudersanam S;
XX
DR WPI; 2001-032161/04.
XX
DR P-PSDB; AAB65621.
XX
PT Nucleic acids encoding kinase polypeptides, useful for diagnosing and
PT treating immune-related diseases and disorders, cardiovascular disease,
PT neurodegenerative diseases and/or cancers -
XX
PS Example 1; Fig 2; 310pp; English.
XX
CC The present sequence encodes a novel protein kinase. The nucleic acids
CC and the protein kinases they encode may be used in the treatment and
CC diagnosis of diseases associated with inappropriate kinase expression
CC such as immune-related diseases and disorders, cardiovascular disease,
CC neurodegenerative diseases and/or cancers. The nucleic acids and
CC complementary sequences may also be used as DNA probes in diagnostic
CC assays. The kinase polypeptides may be used as antigens in the production
CC of antibodies of kinase expression and activity. Anti-kinase antibodies
CC and kinase antagonists may also be used to down regulate kinase
CC expression and activity. Diseases related to kinase expression and
CC activity include rheumatoid arthritis, atherosclerosis, autoimmune
CC disorders, complications of organ transplantation, myocardial infarction,
CC immune disorders, cardiomyopathies, strokes, renal failure,
CC oxidative-stress related disorders, chronic inflammatory bowel disease,
CC chronic inflammatory pelvic disease, multiple sclerosis, asthma,
CC osteoarthritis, psoriasis, rhinitis, autoimmunity, diabetes, cancers and
XX reproductive disorders.
XX
SQ Sequence 2424 BP; 644 A; 575 C; 668 G; 535 T; 2 other;
```

```
Query Match 4.9%; Score 68; DB 22; Length 2424;
Best Local Similarity 57.1%; Pred No. 1.9e-06;
Matches 144; Conservative 0; Mismatches 105; Indels 3; Gaps 1;

QY 558 GAGTTCTTGAAGGCTTCTGCGCTCCAGCCACGTAATTTGCCCGCTCTGTGATCTGTC 617
  || || || || || || || || || || || || || || || || || || || || || ||
Db 249 GAGCTTGTGGAATCGTTCAAATCCAGCGGCCACATACTGCCCTCTCTTCTTCAAGTCTGC 190
QY 618 TAGCTTCCGGATTCCGTGCGGAGTCCGGGGGTGTAGATGTTCTCTACGCGCCCAAGGG 677
  || || || || || || || || || || || || || || || || || || || || || ||
Db 189 CAGGTGGTGACAGGGTGGCCATGACAGGTGTGTAGAGGACCGCACAGCC---AGTGG 133
QY 678 TGCCTGAACGCGCGCGGTACCTCTTCAAGGAAGACTTCAAGCTGACACCTTCTTCTTC 737
  || || || || || || || || || || || || || || || || || || || || || ||
Db 132 GCGCTGCACAGCTGATGTCACCTCGCAGAGGAAGCCCTCATGGTGGGGAAGCGGCTTG 73
QY 738 ATGGATGACGACGCGCGCGCGCGGTAGAGGGTCCCGCTTGGCGGTACACAGCAGCT 797
  || || || || || || || || || || || || || || || || || || || || || ||
Db 72 AGTCACCAACAGCTGGAGCGCTGGGAAGATGGTCCCCATTCGCGTACACCACTACCCT 13
QY 798 CTTCACGACGGG 809
  || || || || || || || || || || || || || || || || || || || || || ||
Db 12 CTGGCTGCTGG 1

RESULT 12
ABN38745
ID ABN38745 standard; DNA; 60 BP.
XX
AC ABN38745;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human spliced transcript detection oligonucleotide SEQ ID NO:11493.
XX
KW Human; mouse; rat; splice transcript; detection; RNA transcript;
KW splice variant; transcriptome; oligonucleotide library; ss.
XX
OS Homo sapiens.
XX
PN WO200210449-A2.
XX
PD 07-FEB-2002.
XX
PF 20-JUL-2001; 2001WO-IB01903.
XX
PR 28-JUL-2000; 2000US-221607P.
XX
PR 02-MAY-2001; 2001US-287724P.
XX
PA (COMP-) COMPUGEN INC.
XX
PI Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
XX
DR WPI; 2002-257383/30.
XX
PT New oligonucleotide libraries comprising oligonucleotides which
PT selectively hybridize to mRNAs transcribed from a transcription unit of
PT a genome, useful for detecting tissue-, pathology-, and
PT developmental-specific genes -
XX
PS Example 1; SEQ ID 11493; 47pp; English.
XX
CC The present invention describes oligonucleotide libraries for detecting
CC messenger RNAs that populate a (sub-)transcriptome. Where the
CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple
CC transcription units that populate a genome. The library comprises
CC several oligonucleotides, each capable of hybridizing selectively to a
CC set of messenger RNAs transcribed from a given transcription unit of
CC the genome, which encodes one or more messenger RNA splice variants.
CC The oligonucleotide libraries are useful for detecting mRNAs from a
CC biological sample, in expression profiling studies, in qualitatively or
CC quantitatively characterising the corresponding transcriptome, and in
CC detecting RNA transcripts and splice variants of human or animal
```

CC transcriptsomes. The libraries may also be used as specialised mini  
 CC libraries to detect transcripts of a sub-transcriptome under a  
 CC particular biological or pathological state, and so allowing the  
 CC detection of tissue- and pathology-specific genes such as those genes  
 CC only expressed in specific tissue under a specific pathological  
 CC condition; to detect developmental specific genes; and to detect RNA  
 CC transcripts and splice variants of a transcriptome of a patient suffering  
 CC from a particular disorder. ABN27253 to ABN59589 represent  
 CC oligonucleotide sequences from rats, humans and mice, which are used in  
 CC the exemplification of the present invention.  
 CC N.B. The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX  
 SQ Sequence 60 BP; 23 A; 15 C; 15 G; 7 T; 0 other;

Query Match 4.3%; Score 60; DB 24; Length 60;  
 Best Local Similarity 100.0%; Pred. No. 3.5e-05;  
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 938 ATCGCTCTCTGAACGAGAACCTGACGAAATCCACAGGTGAAGAGAAAGTAACGGCC 997  
 Db 1 ATCGCTCTCTGAACGAGAACCTGACGAAATCCACAGGTGAAGAGAAAGTAACGGCC 60

## RESULT 13

AAF21831  
 ID AAF21831 standard; DNA; 1645 BP.

AC AAF21831;

XX 27-MAR-2001 (first entry)

DE Human breast and ovarian cancer associated antigen gene SEQ ID 218.

XX Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;  
 KW nototropic; neuroprotective; antiviral; antiallergic; hepatotropic;  
 KW antibacterial; antifungal; antiparasitic; antitumor; anticonvulsant;  
 KW Addison's disease; allergy; autoimmune haemolytic anaemia;  
 KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;  
 KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;  
 KW cardiovascular disorder; wound healing; neurological disease; ds.

OS Homo sapiens.

XX WO200055173-A1.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05881.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX WPI; 2000-611515/58.

DR P-PSDB; AAB58928.

XX New human breast and ovarian cancer associated gene sequences and the  
 PT polypeptides encoded by these genes, useful in the prevention,  
 PT treatment and diagnosis of cancer, immune disorders, cardiovascular  
 PT disorders and neurological diseases -

XX Claim 1; Page 645-646; 1299pp; English.

XX Sequences AAF21614 - AAF22031 represent DNA sequences encoding human  
 CC proteins AAB58711 - AAB59128. The DNA and protein sequences are  
 CC associated with breast and ovarian cancer. Included in the invention are  
 CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the  
 CC isolation and characterisation of the DNA and protein sequences of the

CC invention. The breast and ovarian cancer associated DNA, protein, agonist  
 CC or antagonist sequences exhibit cytostatic; immunosuppressive;  
 CC neuroprotective; antiviral; antiallergic; hepatotropic;  
 CC antibacterial; antifungal; antiparasitic; antitumor; anticonvulsant;  
 CC antibacterial; antifungal; antiparasitic and cardiac activity. The  
 CC polynucleotide and protein sequences are used in the diagnosis of cancer,  
 CC particularly breast and ovarian cancer. The nucleic acid sequences,  
 CC proteins, agonists and antagonists may also be used in the diagnosis,  
 CC prevention and treatment of immune disorders e.g. Addison's disease,  
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,  
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid  
 CC arthritis and ulcerative colitis; cardiovascular disorders such as  
 CC myocardial ischaemias; wound healing; neurological diseases such as  
 CC cerebral anoxia and epilepsy; and infectious diseases.  
 XX  
 SQ Sequence 1645 BP; 400 A; 419 C; 448 G; 374 T; 4 other;

Query Match 3.7%; Score 51.6; DB 21; Length 1645;  
 Best Local Similarity 60.9%; Pred. No. 0.016;  
 Matches 84; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 119 AATTTCTCTCTGTCACAAATATAGTCTCCACACTTGGAAGTGATTTTAAAGAATGTT 178

Db 1469 AATTTCTCTCTGTCACAAATATAGTCTCCACACTTGGAAGTGATTTTAAAGAATGTT 1528

QY 179 TATTTAAATTAAGGATGGATTTCGAAGGAAAAAAATTAAGGAAAGGAAAAAAAC 238

Db 1529 TATTTCTCTCAAAAAAAGGAAAAAAGGAAAAAAGGAAAAAAGGAAAAAAGGAAAA 1588

QY 239 TGAACAGAAACGCAAAA 256

Db 1589 AAAAAAAGGAAAAAAGGAAAAAAGGAAAAAAGGAAAAAAGGAAAAAAGGAAAA 1606

## RESULT 14

AAS45328/c  
 ID AAS45328 standard; DNA; 11735 BP.

AC AAS45328;

XX 18-DEC-2001 (first entry)

DE Chemically pretreated genomic DNA associated with cell cycle #17.

XX Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;  
 KW human immunodeficiency virus; neurodegenerative disorder; solid tumour;  
 KW graft-versus-host disease; glomerular disease; Levey body disease; cancer;  
 KW arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;  
 KW immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;  
 KW PCR primer.

OS Homo sapiens.

PN WO200168911-A2.

XX 20-SEP-2001.

XX 15-MAR-2001; 2001WO-EP02945.

XX 15-MAR-2000; 2000DE-1013847.

PR 06-APR-2000; 2000DE-1019058.

PR 07-APR-2000; 2000DE-1019173.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

PI WPI; 2001-602751/68.

XX Designing primers and probes for analysing diseases associated with  
 PT cytosine methylation state e.g. arthritis, cancer, aging,





**THIS PAGE BLANK (USP10)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 26, 2002, 05:31:33 ; Search time 92,129 Seconds  
(without alignments)  
187,866 Million cell updates/sec

Title: US-09-674-593-2  
Perfect score: 451  
Sequence: 1 MDDDAAPRVEGVAVVHKA.....GAGSPPTNEKLTNPQVKEK 84

Scoring table: BLOSUM62  
Gapop 10.0 , Capext 0.5

Searched: 671580 seqs, 206047115 residues  
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : SPTREMBL\_21.\*
- 1: sp\_archaea.\*
  - 2: sp\_bacteria.\*
  - 3: sp\_fungi.\*
  - 4: sp\_human.\*
  - 5: sp\_invertebrate.\*
  - 6: sp\_mammal.\*
  - 7: sp\_mhc.\*
  - 8: sp\_organelle.\*
  - 9: sp\_phase.\*
  - 10: sp\_plant.\*
  - 11: sp\_prodent.\*
  - 12: sp\_virus.\*
  - 13: sp\_vertebrate.\*
  - 14: sp\_unclassified.\*
  - 15: sp\_virus.\*
  - 16: sp\_bacteriaph.\*
  - 17: sp\_archaeap.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	451	100.0	84	Q9UBP8	Q9ubp8 homo sapien
2	91.5	20.3	1154	Q9P2P0	Q9p2p0 homo sapien
3	91.5	20.3	1431	Q8WWD9	Q8wxd9 homo sapien
4	87	19.3	554	Q9X8S2	Q9x8s2 streptomyce
5	81	18.0	1083	Q86637	Q86637 streptomyce
6	80	17.7	456	Q8WU60	Q8wu60 homo sapien
7	80	17.7	456	Q96DP0	Q96dp0 homo sapien
8	80	17.7	527	Q96RE7	Q96re7 homo sapien
9	79.5	17.6	3247	Q65553	Q65553 bovine herp
10	78.5	17.4	438	Q9C6J4	Q9c6j4 arabidopsis
11	78	17.3	523	Q8RK12	Q8rk12 pseudomonas
12	78	17.3	678	Q8QZT8	Q8qzt8 mus musculu
13	77.5	17.2	1099	Q24096	Q24096 drosophila
14	77.5	17.2	1099	Q24590	Q24590 drosophila
15	77.5	17.2	1105	Q9VA38	Q9va38 drosophila
16	77	17.1	541	Q8RK09	Q8rk09 pseudomonas

17	77	17.1	700	5	P90929	P90929 caenorhabdi
18	77	17.1	1283	5	Q8T5H0	Q8t5h0 anopheles g
19	76.5	17.0	853	4	Q14333	Q14333 homo sapien
20	76	16.9	129	11	Q9CVD0	Q9cvd0 mus musculu
21	75.5	16.7	554	10	Q04123	Q04123 zea mays (m
22	75.5	16.7	843	10	Q94JY7	Q94jy7 arabidopsis
23	75.5	16.7	889	10	Q9FK48	Q9fk48 arabidopsis
24	75.5	16.7	1202	4	Q9P283	Q9p283 homo sapien
25	75.5	16.7	1413	5	Q9VJ38	Q9vj38 drosophila
26	75.5	16.7	1424	5	Q9VJ39	Q9vj39 drosophila
27	75	16.6	539	2	Q9RBW3	Q9rbw3 pseudomonas
28	75	16.6	1122	2	Q85018	Q85018 mycobacteri
29	75	16.6	1151	10	Q93VE9	Q93ve9 oryza sativ
30	75	16.6	1668	5	Q76930	Q76930 drosophila
31	74.5	16.5	298	16	Q8UKS8	Q8uks8 agrobacteri
32	74.5	16.5	3084	12	Q8UZ11	Q8uz11 pseudorabie
33	74	16.4	395	16	Q9RV74	Q9rv74 deinococcus
34	74	16.4	891	10	Q9ZW08	Q9zw08 arabidopsis
35	74	16.4	894	10	Q9FYB2	Q9fyb2 arabidopsis
36	73.5	16.3	251	10	Q9C6X7	Q9c6x7 arabidopsis
37	73.5	16.3	489	10	Q9LNY9	Q9lny9 arabidopsis
38	73.5	16.3	1740	10	Q8S725	Q8s725 oryza sativ
39	73	16.2	1171	5	Q9VHC2	Q9vhc2 drosophila
40	73	16.2	1198	2	Q53730	Q53730 streptomyce
41	72.5	16.1	118	11	Q9CVQ3	Q9cvq3 mus musculu
42	72.5	16.1	336	4	Q16096	Q16096 homo sapien
43	72.5	16.1	490	11	Q8VCL9	Q8vcl9 mus musculu
44	72.5	16.1	491	4	Q9H5S6	Q9h5s6 homo sapien
45	72.5	16.1	491	4	Q8TAJ4	Q8taj4 homo sapien

ALIGNMENTS

RESULT 1	Q9UBP8	PRELIMINARY;	PRT;	84	AA.
ID	Q9UBP8	PRELIMINARY;	PRT;	84	AA.
AC	Q9UBP8;				
DC	Q1-MAY-2000 (TrEMBLrel. 13, Created)				
DT	Q1-MAY-2000 (TrEMBLrel. 13, Last sequence update)				
DT	Q1-MAR-2001 (TrEMBLrel. 16, Last annotation update)				
DE	RU2AS protein.				
GN	RU2.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=20069887; PubMed=10601354;				
RA	Van den Eynde B.J., Gaugler B., Probst-Keppler M., Michaux L.,				
RA	Devuyt O., Lorge F., Weynants P., Boon T.;				
RT	"A new antigen recognized by cytolytic T lymphocytes on a human kidney				
RL	J. Exp. Med. 190:1793-1800(1999).				
DR	EMBL; AF181722; AAF23613.1; -				
DR	ENBL; AF181720; AAF23611.1; -				
SQ	SEQUENCE 84 AA; 8969 MW; 70B739F173A9E560 CRC64;				

Query Match 100.0%; Score 451; DB 4; Length 84;  
Best Local Similarity 100.0%; Pred. No. 2.5e-39;  
Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MDDDAAPRVEGVAVVHKAHDLGRLQVAGVGAGAAAHLPWWPPQLAASRRAPPLSQRP	60
Db	1	MDDDAAPRVEGVAVVHKAHDLGRLQVAGVGAGAAAHLPWWPPQLAASRRAPPLSQRP	60
Qy	61	HRTQAGSGPPTNEKLTNPQVKEK	84
Db	61	HRTQAGSGPPTNEKLTNPQVKEK	84
RESULT 2			

```
Q9P2P0
ID Q9P2P0 PRELIMINARY; PRT; 1154 AA.
AC Q9P2P0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE KIAA1306 protein (Fragment).
GN KIAA1306
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=20181126; PubMed=10718198;
RA Nagase T., Kikuno R., Ishikawa K., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes.XVI.
RT The complete sequences of 150 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 7:65-73(2000).
DR EMBL; AB037727; BAA92544.1; -
DR HSP; P29323; 184F.
DR InterPro; IPR002965; P_Rich_extensn.
DR InterPro; IPR001660; SAM.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00536; SAM; 2.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR01217; PRICHEXTENS.
DR SMART; SM00454; SAM; 2.
DR SMART; SM00326; SH3; 1.
DR SMART; SM00248; ANK; 6.
DR SMART; SM00454; SAM; 2.
DR PROSITE; PS00002; SH3; 1.
DR PROSITE; PS00002; SH3; 1.
FT NON_TER 1
SQ SEQUENCE 1154 AA; 119815 MW; 05D603B5B98616B8 CRC64;

Query Match 20.3%; Score 91.5; DB 4; Length 1154;
Best Local Similarity 37.7%; Pred. No. 0.41;
Matches 23; Conservative 9; Mismatches 20; Indels 9; Gaps 2;

Qy 31 PGAAAHLPWPPLPPGEARKPAKPVLPVTPVVKLGSPPTTSKKVPLPGSPSEV 981
Db 924 PPTDLAHLPLPPPEGEARKPAKPVLPVTPVVKLGSPPTTSKKVPLPGSPSEV 983

Qy 82 K 82
Db 984 K 984

RESULT 3
Q8WXD9
ID Q8WXD9 PRELIMINARY; PRT; 1431 AA.
AC Q8WXD9;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Cask-interacting protein 1.
DR CASKIN1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Tabuchi K., Biederer T., Butz S., Sudhof T.C.;
RT "Cask participates in two alternative tripartite complexes in which
RT Mint 1 and Caskin 1 compete for binding.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF451977; AAL49758.1; -.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000104; Antifreeze_1.
DR InterPro; IPR001660; SAM.
DR InterPro; IPR001452; SH3.
DR InterPro; IPR003993; treacle.

Q9P2P0
ID Q9P2P0 PRELIMINARY; PRT; 1154 AA.
AC Q9P2P0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE KIAA1306 protein (Fragment).
GN KIAA1306
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=20181126; PubMed=10718198;
RA Nagase T., Kikuno R., Ishikawa K., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes.XVI.
RT The complete sequences of 150 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 7:65-73(2000).
DR EMBL; AB037727; BAA92544.1; -
DR HSP; P29323; 184F.
DR InterPro; IPR002965; P_Rich_extensn.
DR InterPro; IPR001660; SAM.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00536; SAM; 2.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR01217; PRICHEXTENS.
DR SMART; SM00454; SAM; 2.
DR SMART; SM00326; SH3; 1.
DR SMART; SM00248; ANK; 6.
DR SMART; SM00454; SAM; 2.
DR PROSITE; PS00002; SH3; 1.
DR PROSITE; PS00002; SH3; 1.
FT NON_TER 1
SQ SEQUENCE 1154 AA; 119815 MW; 05D603B5B98616B8 CRC64;

Query Match 20.3%; Score 91.5; DB 4; Length 1154;
Best Local Similarity 37.7%; Pred. No. 0.41;
Matches 23; Conservative 9; Mismatches 20; Indels 9; Gaps 2;

Qy 31 PGAAAHLPWPPLPPGEARKPAKPVLPVTPVVKLGSPPTTSKKVPLPGSPSEV 981
Db 924 PPTDLAHLPLPPPEGEARKPAKPVLPVTPVVKLGSPPTTSKKVPLPGSPSEV 983

Qy 82 K 82
Db 984 K 984

RESULT 3
Q8WXD9
ID Q8WXD9 PRELIMINARY; PRT; 1431 AA.
AC Q8WXD9;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Cask-interacting protein 1.
DR CASKIN1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Tabuchi K., Biederer T., Butz S., Sudhof T.C.;
RT "Cask participates in two alternative tripartite complexes in which
RT Mint 1 and Caskin 1 compete for binding.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF451977; AAL49758.1; -.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000104; Antifreeze_1.
DR InterPro; IPR001660; SAM.
DR InterPro; IPR001452; SH3.
DR InterPro; IPR003993; treacle.
```

```
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00536; SAM; 2.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR01415; ANKYRIN.
DR PRINTS; PR00308; ANTIFREEZE1.
DR PRINTS; PR01503; TREACLE.
DR SMART; SM00248; ANK; 6.
DR SMART; SM00454; SAM; 2.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00088; ANK_REPEAT; 6.
DR PROSITE; PS0297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS0105; SAM_DOMAIN; 2.
DR PROSITE; PS00002; SH3; 1.
SQ SEQUENCE 1431 AA; 149813 MW; CDF60E68B81E884A CRC64;

Query Match 20.3%; Score 91.5; DB 4; Length 1431;
Best Local Similarity 37.7%; Pred. No. 0.51;
Matches 23; Conservative 9; Mismatches 20; Indels 9; Gaps 2;

Qy 31 PGAAAHLPWPPLPPGEARKPAKPVLPVTPVVKLGSPPTTSKKVPLPGSPSEV 1260
Db 1201 PPTDLAHLPLPPPEGEARKPAKPVLPVTPVVKLGSPPTTSKKVPLPGSPSEV 1260

Qy 82 K 82
Db 1261 K 1261

RESULT 4
Q9X8S2
ID Q9X8S2 PRELIMINARY; PRT; 554 AA.
AC Q9X8S2;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative spermidine synthase.
GN SC03655 OR SCH10.33C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Murphy L., Harris D.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapite D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Gobie A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser L., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
```

[illegible]

RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,  
RA Whites O., Alonso J., Altai H., Araujo R., Bowman C.L., Brooks S.Y.,  
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W., Dewar K.,  
RA Dunn M.N., Conn L., Conway A.B., Conway A.R., Cressey T.H., Dewar K.,  
RA Dunn P., Eguu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,  
RA Gail J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,  
RA Hunter J.-L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,

RA	Kim C.J., Koo H.L., Kremenetskaya I., Kurtz D.B., Kwan A., Lam B.,
RA	Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA	Liu X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marzilli A.,
RA	Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA	Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA	Sakano H., Salzer S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA	Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA	Utterback T., Van Aken S., Vaysberg M., Vysotskaya V.S., Walker M.,
RA	Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT	*Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT	thaliana.*;
RL	Nature 408:816-820(2000).
DR	EMBL: AC079284; AAG50940.1; -;
DR	InterPro: IPR004252; Transposase_24.
DR	Pfam: PF03004; Transposase_24; 1.
KW	Hypothetical protein.
QY	SEQUENCE 438 AA; 49165 MW; 1ECB70C57989CA22 CRC64;
Query Match	17.4%; Score 78.5; DB 10; Length 438;
Best Local Similarity	32.8%; Pred. No. 3.3;
Matches	21; Conservative 9; Mismatches 23; Indels 11; Gaps
QY	27 QVAGPGAAAHLPWPWP-QL-----AASRRRAPLSQRPHRTQAGSPPTNEKLTNP 79
Db	53 QQHQGSTSAPHVRNYPPLPQLFONSTTQHQHAPPLSEVHYTPPTAPPOQ----TPP 108
QY	80 QVKE 83
Db	109 QSRD 112
RESULT 11	
ID	Q8RK12 PRELIMINARY; PRT; 523 AA.
AC	Q8RK12;
DT	01-JUN-2002 (TrEMBLrel. 21, Created)
DT	01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE	VirPphApgY protein.
OS	VirPphApgY.
GN	Pseudomonas syringae (pv. glycinea).
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC	Pseudomonas.
OX	NCBI_TaxID=318;
RP	[1]
RP	SEQUENCE FROM N.A.
RC	SPRAIN=49A/90;
RA	Jackson R.W., Mansfield J.W., Ammoun H., Dutton L.C., Wharton B.,
RA	Ortiz-Barredo A., Arnold D.L., Tsiamis G., Sesna A., Butcher D.,
RA	Boch J., Kim Y.J., Martin G.B., Tegli S., Murillo J., Vivian A.;
RT	*Pathogenicity islands in plant pathogenic pseudomonads: location and
RT	activity of virPphA homologues from pathovars of Pseudomonas syringae
RT	and P. savastanoi*;
RL	Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR	EMBL: AJ439728; CAD29299.1; -;
QY	SEQUENCE 523 AA; 57996 MW; 2F3264B6C4BA1374 CRC64;
Query Match	17.3%; Score 78; DB 2; Length 523;
Best Local Similarity	28.1%; Pred. No. 4.5;
Matches	25; Conservative 12; Mismatches 28; Indels 24; Gaps
QY	8 RVEGVVPYVHKHALHDLGRVAGPGAAAAHLRPWPP-----QLAASRRAP 54
Db	17 RTDGEPTVERE---HDSRR-----SASSANSPEFPFPAPAESGRORLLRSSALSROTRE 68
QY	55 PLSQRPHRTQAGSPPTNEKLTNPQVKE 83
Db	69 WLEATPARVQCATPPAEARO---SPEAQ 94
RESULT 12	
Q8QZT8	PRELIMINARY; PRT; 678 AA.
ID	Q8QZT8

```

Qy 17 HKHALH-----DGLRQVAGP-----GAAAHLPWPWPQA 47
Db 162 HSHHTQSSRTVGNPGNGGFSFSEVAPPAPPRNPTASSAATPPPPVPTSOA 221

Qy 48 ASRREAPPLSQRPHR-----TOGAGSPPETNEKLTNPQ 80
Db 222 YVKRRSPALNNRPPAIAPTQRGNSPVITONGLNKPQ 258

RESULT 14
Q24590 PRELIMINARY; PRT: 1099 AA.
ID Q24590;
AC Q24590;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE Tumor suppressor.
GN WTS OR CG12072.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95212904; PubMed=7698644;
RA Justice R.W., Zilian O., Woods D.F., Noll M., Bryant P.J.;
RT "The Drosophila tumor suppressor gene warts encodes a homolog of human
RT myotonic dystrophy kinase and is required for the control of cell
RT shape and proliferation.";
RL Genes Dev. 9:534-546(1995).
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL: L39837; AA73959.1;
DR HSSP; P05132; 1APM.
DR FlyBase; FBgn0011739; wts.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR000961; pkinase_C.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00433; pkinase_C; 1.
DR ProDom; PD000001; Euk_pkinase; 2.
DR SMART; SM00220; S_TK; 1.
DR SMART; SM01133; S_TK_X; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 1099 AA; 121633 MW; EB35192A10470805 CRC64;

Query Match 17.2%; Score 77.5; DB 5; Length 1099;
Best Local Similarity 27.8%; Pred. No. 11;
Matches 27; Conservative 5; Mismatches 32; Indels 33; Gaps 3;

Qy 17 HKHALH-----DGLRQVAGP-----GAAAHLPWPWPQA 47
Db 162 HSHHTQSSRTVGNPGNGGFSFSEVAPPAPPRNPTASSAATPPPPVPTSOA 221

Qy 48 ASRREAPPLSQRPHR-----TOGAGSPPETNEKLTNPQ 80
Db 222 YVKRRSPALNNRPPAIAPTQRGNSPVITONGLNKPQ 258

RESULT 15
Q9VA38 PRELIMINARY; PRT: 1105 AA.
ID Q9VA38;
AC Q9VA38;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE WTS protein.
GN WTS OR CG12072.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

```

```

OC pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BERKELEY;
RA MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.C., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balles R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottker P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
SC Science 287:2185-2195(2000).
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; AE003775; AAF57085.1;
DR HSSP; P05132; 1APM.
DR FlyBase; FBgn0011739; wts.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR000961; pkinase_C.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00433; pkinase_C; 1.
DR ProDom; PD000001; Euk_pkinase; 2.
DR SMART; SM00220; S_TK; 1.
DR SMART; SM01133; S_TK_X; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 1105 AA; 122488 MW; 7D09D7AA9BA1A856 CRC64;

Query Match 17.2%; Score 77.5; DB 5; Length 1105;
Best Local Similarity 27.8%; Pred. No. 11;
Matches 27; Conservative 5; Mismatches 32; Indels 33; Gaps 3;

Qy 17 HKHALH-----DGLRQVAGP-----GAAAHLPWPWPQA 47
Db 162 HSHHTQSSRTVGNPGNGGFSFSEVAPPAPPRNPTASSAATPPPPVPTSOA 221

Qy 48 ASRREAPPLSQRPHR-----TOGAGSPPETNEKLTNPQ 80

```



Db 222 YVRRSPALNNRPPAIAPTQRGNSPVITQNGLKNPQ 258

Search completed: November 26, 2002, 17:59:16  
Job time : 107.129 secs

**THIS PAGE BLANK (USPTO)**



A;Cross-references: GB:AE005173; NID:gl1094698; PIDN:AAG29634.1; GSPDB:GN0014141  
C;Genetics:  
A;Gene: F8A12.27  
A;Map position: 1

Query Match 17.4%; Score 78.5; DB 2; Length 438;  
Best Local Similarity 32.8%; Pred. No. 3.5;  
Matches 21; Conservative 9; Mismatches 23; Indels 11; Gaps 3;

27 QVAGPAAAAHLPRWPPP-QL-----AASREAPPLSQRPHTQAGSPPETNEKLTNP 79  
 53 OCHGPSTSAHVRYNPPLOLFONSTPOHOAPPLSEEVHYTPPTAPPOQ-----TTP 108

Qy	80 QVKE 83
	↓ ::
Db	109 OSRD 112

RESULT 3  
S44759  
C14B9.5 protein - *Caenorhabditis elegans*  
C:Species: *caenorhabditis elegans*  
C:Date: 14-Sep-1994 #sequence\_revision 12-May-1995 #text\_change 09-Sep-1997  
C:Accession: S44759  
R:Favello, A. D.  
submitted to the EMBL Data Library, May 1993  
A:Description: Sequence of the *C. elegans* cosmid C14B9.

Query Match	17.3%	Score 78;	DB 2;	Length 789;
Best Local Similarity	29.5%;	Pred. No. 7;		
Matches	26:	Conservative	13:	Mismatches
			25:	Indels
			24:	Gaps
			4:	

Qy	9	VEGVPVAVHKHALHDGLRQVAGPGAAAHLPWP	PPQLAASRREAPPLSQRP	HR-----T	63
Dh	331	MOGAP---HPHT-----OLAGPSSSR	TSGRGVSPAL	SASRSVAP	TGGAAGRSITAGT

```

QY      64  QGAGSPPETNEKL-----TNPQV 81
      | | : | : |
Dh    382  SCHSSSTDYKNEFLAERMNVEFKPTHPOL 408

```

RESULT 4  
A56155  
tumor suppressor protein warts (EC 2.7.1.-) - fruit fly (*Drosophila melanogaster*)  
C:Species: *Drosophila melanogaster*  
C:Date: 03-Oct-1995 #sequence\_revision 03-Oct-1995 #text\_change 24-Sep-1999  
C:Accession: A56155  
R:Justice, R.W.; Zillian, O.; Woods, D.F.; Noll, M.; Bryant, P.J.  
Genes Dev. 9, 534-546, 1995  
A:Title: The *Drosophila* tumor suppressor gene warts encodes a homolog of human myotonic  
A:Reference number: A56155; MUID:95212904; PMID:7698644

Query Match

Best Local Similarity	27.8%;	Pred. No. 11;	
Matches	27;	Conservative	32;
		Mismatches	Indels
			33;
		Gaps	3;

Qy 17 HKHALH-----DGLRQVAGP-----GAAAAHLPRWPPQLA 47

Dh 163 HSHHHHPSSRTVGNPGNGGFSPPSGFSEVAPPAPPPRNPTASSAA"PPPPVPPTSOA 221

Qy 48 ASRREAPPLSQPHR---TQAGSGPPETNEKLTNPQ 80

RESULT 5  
S40887  
RVSI67 protein - yeast (*Saccharomyces cerevisiae*)  
N:Alternate names: protein YDR388w  
C:Species: *Saccharomyces cerevisiae*  
C:Date: 31-Mar-1992 #sequence\_revision 06-Feb-1995 #text\_change 21-Jul-2000  
C:Accession: S40887; S69672  
R:Bauer, F.; Urdaci, M.; Aigle, M.; Crouzet, M.  
Mol. Cell. Biol. 13, 5070-5084, 1993  
A:Title: Alteration of a yeast SH3 protein leads to conditional viability with defect  
A:Reference number: S40887; PMID:9330299; PMID:8336735  
A:Accession: S40887  
A:Molecule type: DNA  
A:Residues: 1-482 <BAU>  
A:Cross-references: EMBL:M92092; NID:gl72615; PID:AAA35051.1; PID:gl72616  
R:Dietrich, F.S.

Query Match 17.1%; Score 77; DB 1; Length 482;  
Best Local Similarity 34.1%; Pred. No. 5.3;  
Matches 28; Conservative 8; Mismatches 34; Indels 12; Gaps 4;

QY 5 AAPRVEGVPVAVHKHALHDLGRQVAG--PGAAAAHLPRWPPQLAASRRAPPLS----- 57  
 ||| ||| : ||| : ||| : ||| : ||| :  
 332 AAPDVAAGDPAAGVCTAAAVCTAAAGAAAGAVGCTYPOVAAA---OSPLTGLGFG 388

QY 58 QRPHTQGAGSPPETNEKLTNP 79  
| | : | | | : | | : |  
388 CSP0000--GPPPAYSNPI.TSP 408

RESULT 6  
T20550  
hypothetical protein F07C6.4b - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T20550; T23678  
R:Steward, C.  
submitted to the EMBL Data Library, February 1996  
A:Reference number: Z19290  
A:Accession: T20550  
A>Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-700 <WILL>  
A:Cross-References: EMBL:269859; PIDN:CAA93486.1; GSPDB:GN00022; CESP:F07C6.4b  
A:Experimental source: clone F07C6

```
R;Yu, S.; Fiss, E.; Jacobs Jr., W.R.  
J. Bacteriol. 180, 4676-4685, 1998  
A:title: Analysis of the exochelin locus in mycobacterium smegmatis: biosynthesis gen  
A:Reference number: Z17898; MUID:98389687; PMID:9721311  
A:Accession: Tl14163  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 534-1122 <YUS>  
A:Cross-references: EMBL:AF027770; NID:g3560502; PID:g3560505; PIDN:AACB2548.1  
C:Genetics:  
A:Gene: exit  
C:Superfamily: Mycobacterium tuberculosis probable ABC transporter Rv0194; ATP-bindin
```

	Query Match	16.6%; Score 75;	DB 2; Length 1122;
	Best Local Similarity	31.0%;	Pred. No. 19;
Matches	22; Conservative	9; Mismatches 28;	Indels 12; Gaps 3;

```
QY      15 AVVKKHALHD--GLRQAVGP-----AAAAHLPRWPPPLQAASREAPLSQRPHRTGAGS 68  
         |   |   |   | :|||       :    |||   :|||   |||  
Db     479 AAHARTHGNGGHQQSGPARRGPGAARH-----RASGORRHTPFSGRRLLRLPGSGD 532  
                                         ::::|::|::|::|  
  
QY      69 PPETNEKLTPP 79  
         |   |:|   |   |  
Db     533 PMISDEPVTPP 543  
         |   |:|   |   |
```

RESULT 9

Tl13748

sex comb protein - fruit fly (*Drosophila melanogaster*)

C:Species: *Drosophila melanogaster*

C:date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #tex\_change 17-Nov-2000

C:Accession: Tl13748

R:Sincclair, D.A.R.; Milne, T.A.; Hodgson, J.W.; Shellard, J.; Salinas, C.A.; Kyba, M.

A:title: The Additional sex combs gene of *Drosophila* encodes a chromatin protein that

A:Reference number: Z17750; MUID:98146384; PMID:9477319

A:Accession: Tl13748

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-1668 <SIN>

A:Cross-references: EMBL:AJ001164; NID:g3292938; PIDN:CAA04568.1; PID:g3292939

C:Genetics:

A:Cross-references: FlyBase:FBgn0000142

C:Function:

A>Description: Involved in repression of homeotic loci

	Query Match	16.6%; Score 75;	DB 2; Length 1668;
	Best Local Similarity	35.0%;	Pred. No. 29;
Matches	21; Conservative	4; Mismatches 11;	Indels 24; Gaps 3;

```
QY      13 PVAVKHALHDGLGRQAVGGAAAHLPWPVPPO-----LAASFKEAPPILSQRP-HRT 63  
         |||::|   |||   |||   |||   |||   |||   |||   |||   |||   |||  
Db     676 PVLRHR-----PAHHPHLPPPQHCLRAPVAPAKHKRPPQLHRRPEHRT 720  
                                     ::::|::|::|::|::|::|::|::|::|::|::|
```

RESULT 10

AD3162

conserved hypothetical protein Atu5021 [imported] - Agrobacterium tumefaciens (strain)

C:Species: Agrobacterium tumefaciens

C:date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #tex\_change 11-Jan-2002

C:Accession: AD3162

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo

erage, G.; Gillet, W.B.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McCl

i; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam

ster, E.W.

A:title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:reference number: AB2577; PMID:11743193

A:Accession: AD3162

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-298 <KUR>





**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2002, 17:06:07; Search time 88.5161 Seconds  
(without alignments)  
126.452 Million cell updates/sec

Title: US-09-674-593-2

Perfect score: 451

Sequence: 1 MDDDAAPRVGVVAVHKA.....GAGSPETNEKLTNPQVKEK 84

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

A\_Geneseq\_101002:\*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*

2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*

3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*

4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*

5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*

6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*

7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*

8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*

9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*

10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*

11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*

12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*

13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*

14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*

15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*

16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*

17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*

18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*

19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*

20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*

21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*

22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*

23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	451	100.0	84	21	AAV53809
2	86.5	19.2	1352	23	AAE18212
3	82.5	18.3	128	22	AAO11577
4	82	18.2	201	22	ABG03489
5	82	18.2	806	22	ABG04954
6	81.5	18.1	408	20	AA41111
7	81.5	18.1	430	20	AA41110
8	81.5	18.1	430	21	AA410785
9	81.5	18.1	430	22	AAU09901
10	81.5	18.1	430	22	AAE05517

11	81.5	18.1	430	22	AAE95627	Human protein sequ
12	81.5	18.1	878	22	AAE02358	Human tumour necro
13	80	17.7	479	22	AAO13891	Human polypeptide
14	78.5	17.4	1078	22	ABG22869	Novel human diagno
15	77.5	17.2	160	22	ABG13164	Novel human diagno
16	77.5	17.2	1099	17	AAW05177	Lats large tumour
17	77.5	17.2	1099	21	AAV70393	Drosophila melanog
18	77.5	17.2	1105	22	ABE58401	Drosophila melanog
19	77	17.1	212	22	AAO12755	Human polypeptide
20	77	17.1	526	22	ABG20645	Novel human diagno
21	76.5	17.0	853	22	ABG12054	Novel human diagno
22	76.5	17.0	1072	22	ABG04157	Novel human diagno
23	76	16.9	111	21	AAE26081	zea mays protein f
24	76	16.9	621	22	AAU33270	Novel human secret
25	76	16.9	624	22	AAW42014	Human polypeptide
26	76	16.9	625	22	ABG21889	Novel human diagno
27	76	16.9	655	22	ABG19250	Novel human diagno
28	75.5	16.7	72	22	ABG13874	Novel human diagno
29	75.5	16.7	288	22	ABG21768	Novel human diagno
30	75.5	16.7	457	22	ABE11341	Human type II proc
31	75.5	16.7	1413	22	ABE60857	Drosophila melanog
32	75.5	16.7	1424	22	ABE60854	Drosophila melanog
33	75	16.6	510	22	ABG22923	Novel human diagno
34	75	16.6	906	22	AAE92916	Human protein sequ
35	74.5	16.5	120	22	AAO13454	Human polypeptide
36	74.5	16.5	174	22	ABG23384	Novel human diagno
37	73.5	16.3	202	22	ABG29088	Novel human diagno
38	73.5	16.3	700	11	AAE03663	Human myb related
39	73	16.2	93	22	AAU39584	Propionibacterium
40	73	16.2	138	22	AAU41659	Propionibacterium
41	73	16.2	1171	22	ABE64387	Drosophila melanog
42	72.5	16.1	193	22	ABE31899	Peptide #4550 enco
43	72.5	16.1	193	22	ABE37137	Peptide #4643 enco
44	72.5	16.1	193	22	ABE22450	Protein #4449 enco
45	72.5	16.1	193	22	AAE57857	Human brain expres

#### ALIGNMENTS

RESULT 1  
AAV53809  
ID AAV53809 standard; Protein: 84 AA.  
XX  
AC AAV53809;  
XX  
DT 22-FEB-2000 (first entry)  
XX  
DE Protein encoded by tmour rejection antigen RUR-1 antisense cDNA.  
XX  
KW Antisense; human; tumour rejection antigen: RUR-1; tumour; cancer;  
KW renal cell carcinoma; colorectal carcinoma; melanoma; sarcoma;  
KW leukaemia.  
XX  
OS Homo sapiens.  
OS  
PN WO9558546-A1.  
PN  
PD 18-NOV-1999.  
PD  
PF 13-MAY-1999; 99WO-US10424.  
PF  
XX 13-MAY-1998; 98US-0085318.  
XX  
XX (LUDW-) LUDWIG INST CANCER RES.  
XX  
XX Van Den Eynde B, Boon-Falleur T;  
XX  
XX WPI; 2000-053076/04.  
XX  
XX N-PSDB; AAZ36643.  
XX  
XX New isolated tumour rejection antigen RUR-1 nucleic acids, used for,  
XX e.g. treatment of cancers -

XX  
PS Claim 16; Fig 5; 75pp; English.  
XX  
CC The present sequence is encoded by the antisense cDNA sequence of human  
CC tumour rejection antigen RUR-1. This cDNA sequence is the antisense  
CC strand of a ubiquitously expressed gene. The antisense strand codes for  
CC a polypeptide which is preferentially expressed in tumour samples and  
CC a tumour-derived cell lines. The polypeptide is unrelated to any TRAP  
CC protein. The cDNA sequence was isolated from a renal cell carcinoma line  
CC LB9211-RC. The RUR-1 nucleic acids and polypeptides can be used for  
CC diagnosis, prognosis or treatment of a disorder characterized by the  
CC expression of a RUR-1 antisense cDNA molecule or an expression product,  
CC such as cancers, e.g. renal cell carcinoma, colorectal carcinoma,  
CC melanoma, sarcoma or leukaemia.  
XX  
SQ Sequence 84 AA;

Query Match 100.0%; Score 451; DB 21; Length 84;  
Best Local Similarity 100.0%; Pred. No. 4.1e-41;  
Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 MDDAARVEGVAVVHKHALDGLRQVAGPGAAAHLPWPPLQLAASRRAPPLSORP 60  
Db 1 MDDAARVEGVAVVHKHALDGLRQVAGPGAAAHLPWPPLQLAASRRAPPLSORP 60  
Oy 61 HRTQAGSPPETNEKLTNPQVKEK 84  
Db 61 HRTQAGSPPETNEKLTNPQVKEK 84

RESULT 2  
AAE18212  
ID AAE18212 standard; Protein; 1352 AA.  
XX  
AC AAE18212;  
XX  
DT 07-MAY-2002 (first entry)  
XX  
DE Human MOL4 protein.  
XX  
KW Secreted molecule; MOL4 protein; MOLX; cardiomyopathy; atherosclerosis;  
KW diabetes; chromosomal disorder; albinism; anaemia; psoriasis; scarring;  
KW liver cirrhosis; neurodegeneration; osteoarthritis; organ rejection;  
KW cerebral thrombosis; hypertension; systemic lupus erythematosus; asthma;  
KW immune disease; ischaemia; immunodeficiency; rheumatoid arthritis; ulcer;  
KW HIV; human immuno deficiency virus; hepatitis; haemostatic disease; pain;  
KW haematopoietic; thrombolytic; thrombocytopaenia; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease; muscular disease; stress;  
KW ocular disease; growth disorder; depression; epilepsy; contraceptive;  
KW vulneryary; osteopathic; haemostatic; tranquiliser; antidepressant;  
KW analgesic; vasotropic; hypotensive; gene therapy.  
XX  
OS Homo sapiens.  
XX  
XX WO2000206339-A2.  
XX  
XX 24-JAN-2002.  
XX  
XX 03-JUL-2001; 2001WO-US21249.  
XX  
XX 03-JUL-2000; 2000US-215854P.  
XX  
XX 03-JUL-2000; 2000US-215856P.  
XX  
XX 03-JUL-2000; 2000US-215902P.  
XX  
XX 07-JUL-2000; 2000US-216585P.  
XX  
XX 07-JUL-2000; 2000US-216586P.  
XX  
XX 07-JUL-2000; 2000US-216722P.  
XX  
XX 17-JUL-2000; 2000US-218622P.  
XX  
XX 27-JUL-2000; 2000US-218992P.  
XX  
XX 14-FEB-2001; 2001US-268734P.  
XX  
XX (CURA-) CURAGEN CORP.

PI Spaderna SK, Tchernev V, Liu X, Shenoy S, Spylak K, Zerhusen B;  
PI Patturajan M, Taupier RJ, Rastelli L, Grosse WM, Szekeres ES;  
PI Alsobrook J, Lepley DM, Shen L, Burgess CE, Shimkets RA;  
PI Padigaru M;  
XX  
DR WPI; 2002-155038/20.  
DR N-PSDB; AAD28946.  
XX  
XX Nucleic acids encoding secreted polypeptides, designated MOLX  
PT polypeptides, useful for treating a MOLX-associated disorder, e.g.  
PT cardiomyopathy, atherosclerosis, diabetes and metabolic disorders -  
XX  
XX Claim 1; Page 50; 223pp; English.  
XX  
XX The patent discloses nucleic acid sequences encoding novel secreted  
CC molecule (MOLX) polypeptides, designated MOLX polypeptides (i.e. a MOL  
CC protein where X is an integer from 1 to 8). Sequences of the invention  
CC are useful for treating or preventing a MOLX-associated disorder in  
CC humans. They are useful for treating or preventing cardiomyopathy,  
CC atherosclerosis and disorders related to cell signal processing and  
CC metabolic pathway modulation. The MOLX antibodies are useful for  
CC treating or preventing diabetes and disorders related to cell signal  
CC processing and metabolic pathway modulation. MOLX sequences are useful  
CC for the treatment or diagnosis of other MOLX-associated disorders, e.g.  
CC chromosomal disorders, albinism, anaemia, liver cirrhosis, psoriasis,  
CC scarring, neurodegeneration, osteoarthritis, organ rejection, cerebral  
CC thrombosis, ischaemia, hypertension, systemic lupus erythematosus, immune  
CC diseases, immunodeficiency, HIV (human immuno deficiency virus), viral,  
CC bacterial and fungal infections, hepatitis, rheumatoid arthritis, asthma,  
CC haematopoietic, thrombolytic, haemostatic diseases, thrombocytopaenia,  
CC ulcers, Alzheimer's disease, Parkinson's disease, Huntington's disease,  
CC ocular disease, muscular diseases, growth disorders, loss of libido,  
CC stress, depression, pain and epilepsy. They are useful for preventing  
CC chemotherapy side effects and as contraceptives. Sequences of the  
CC invention are also useful for gene therapy. The present sequence  
XX is human semaphorin-like protein, MOL4.  
XX  
SQ Sequence 1352 AA;

Query Match 19.2%; Score 86.5; DB 23; Length 1352;  
Best Local Similarity 39.7%; Pred. No. 1.3;  
Matches 27; Conservative 3; Mismatches 35; Indels 3; Gaps 2;

Oy 5 AARVEGVAVVHKHALHD--GLRQVAGPGAAAHLPWPPLQLAASRRAPPLSORPHR 62  
Db 120 AHPAAAGSPSAAGAGPGDCTGALRAGSCAAAPFPDRPPAHLVSSRSAPPGRPRG 179  
Oy 63 TQAGSGSP 70  
Db 180 T-GHLHPP 186

RESULT 3  
AAO11577  
ID AAO11577 standard; Protein; 128 AA.  
XX  
XX AAO11577;  
XX  
XX 06-NOV-2001 (first entry)  
XX  
XX Human polypeptide SEQ ID NO 25469.  
XX  
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
XX tissue growth factor; immunomodulatory; cancer; leukaemia;  
XX nervous system disorders; arthritis; inflammation.  
XX  
XX Homo sapiens.  
XX  
XX WO200164835-A2.  
XX  
XX 07-SEP-2001.  
XX







appetite; memory; stress; food additive.

Homo sapiens.

Key Peptide Location/Qualifiers

1..25

/label= Signal\_peptide

/note= "Predicted leader sequence for TNF related gene 12"

Protein 26..430

/label= Mature\_human\_TNF\_related\_gene\_12

/note= "Tumour necrosis factor related gene 12"

Domain 26..164

/note= "Predicted extracellular domain"

Domain 48..71

/note= "Predicted Cysteine rich domain"

Domain 165..181

/note= "Predicted transmembrane domain"

Domain 182..430

/note= "Predicted intracellular domain"

WO200181402-A1.

01-NOV-2001.

19-APR-2001; 2001WO-US12762.

19-APR-2000; 2000US-198388P.

18-APR-2001; 2001US-0836607.

(HUMA-) HUMAN GENOME SCI INC.

Ruben SM, Ni J;

WPI: 2001-616772/71.

N-PSDB; AAS15031.

Nucleic acid molecules encoding human TNFR Related Gene 12 proteins, used in preventing, treating or ameliorating a disorder, e.g. Alzheimer's and Parkinson's diseases and cancers -

Claim 11; Fig 1A-C; 272pp; English.

The invention describes isolated nucleic acid molecules encoding Human Tumour necrosis factor receptor related gene 12 (I) proteins. These are used: in gene therapy (antisense) and vaccines; to prevent, or treat diseases associated with inappropriate expression of (I). The nucleic acids and complementary sequences may be used as DNA probes in diagnostic assays to detect and quantitate the presence of similar nucleic acid sequences in samples, and identify patients needing restorative therapy. The (I) proteins are useful antigens in antibody production e.g. for detecting (I) in samples, and in assays to identify modulators of the expression and activity of the (I) proteins. The anti-(I) antibodies and antagonists may be used to down regulate expression and activity of the proteins. The disorders include: immune/autoimmune diseases (e.g. anaemia and rheumatoid arthritis), neurological diseases (e.g. Alzheimer's disease and Parkinson's disease) cancers and hyperproliferative disorders (e.g. neoplasms of the breast or liver, Sezary syndrome and Gaucher's disease), cardio-/cerebrovascular disorders (e.g. cardiac arrest, tachycardia and thrombosis), infections caused by bacteria, viruses and fungi and ocular disorders (e.g. corneal infections). The agonists, antagonists, antibodies, polypeptides and polynucleotides can also be used to promote wound healing, maintain organs before transplantation, support cell culture of primary tissues, induce mesodermal tissue to differentiate in embryos, modulate differentiation of embryonic stem cells, modulate mammalian characteristics (e.g. height and weight), modulate e.g. catabolism, anabolism, mental state, biorhythms, cardiac rhythms, reproductive potential, appetite, memory and stress. The polypeptides can also be used as an additive to increase or decrease storage capabilities and nutritional content of food. This sequence is the human tumour necrosis factor (TNF) related gene 12 described in the method of the invention.

SQ Sequence 430 AA;

Query Match 18.1%; Score 81.5; DB 22; Length 430;

Best Local Similarity 28.2%; Pred. No. 1.3;

Matches 22; Conservative 11; Mismatches 38; Indels 7; Gaps 3;

Qy 7 PRVEGVP-VAVHKHALH--DGLRQVAGPCGAAAHLPWP----PPQLAASREAPPLSOR 59

Db 266 PAPPNVPHICPHRHHLHTVQGLASLSPCCSCRSQKKPEVLLSPFAVAATTPVPSLLPN 325

Qy 60 PHRTQAGSPETNEKLT 77

Db 326 PTRVPKAGAKAGQGEIT 343

RESULT 10

AAE05517

ID AAE05517 standard; Protein: 430 AA.

XX AAE05517;

XX 24-SEP-2001 (first entry)

XX Human tumour necrosis factor receptor/osteoprotegerin-like protein #1.

XX Human; cytostatic; tumour necrosis factor receptor/osteoprotegerin;

XX TNFR/OPG-like protein; antisense inhibitor; gene therapy.

XX Homo sapiens.

XX Key Location/Qualifiers

XX FH Peptide 1..25

XX FT /label= Signal\_peptide

XX FT Protein 26..430

XX FT /label= Mature\_TNFR/OPG-like\_protein

XX FT Domain 1..162

XX FT /label= Extracellular\_domain

XX FT Domain 163..183

XX FT /label= Transmembrane\_domain

XX WO200144472-A1.

XX 21-JUN-2001.

XX 14-DEC-2000; 2000WO-US33858.

XX 16-DEC-1999; 99US-0172306.

XX (AMGE-) AMGEN INC.

XX Jing S, Welcher AA, Fox GM, Shu J, Boedigheimer MJ, Bennett BD;

XX WPI: 2001-451665/48.

XX N-PSDB; AAD10353.

XX New tumor necrosis factor receptor / osteoprotegerin-like

XX (TNFR/OPG-like) polypeptides useful for diagnosis and treatment of

XX associated disease -

XX Claim 14; Fig 3; 208pp; English.

XX The present sequence is tumour necrosis factor receptor/osteoprotegerin

XX (TNFR/OPG)-like protein from human. TNFR/OPG-like protein is

XX useful for the treatment, prevention or amelioration of a medical

XX condition in a mammal resulting from decreased levels of TNFR/OPG

XX protein. Nucleic acid sequences of the present invention are used to

XX map locations of the TNFR/OPG genes and related genes. They are also

XX used as antisense inhibitors of TNFR/OPG expression. TNFR/OPG-like

XX protein and cDNA sequences are also used in gene therapy.

XX SQ Sequence 430 AA;

Query Match 18.1%; Score 81.5; DB 22; Length 430;









**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 25, 2002, 18:45:47 : Search time 27.0968 Seconds  
(without alignments)  
128.577 Million cell updates/sec

Title: US-09-674-593-2

Perfect score: 451

Sequence: 1 MDDDAAPRVEGVPAVHKHA.....GAGSPETNEKLTNPQVKEK 84

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result. No.	Query	Score	Length	ID	Description
1	82.5	18.3	430	1	TRLT_MACFA
2	81.5	18.1	430	1	TRLT_HUMAN
3	78	17.3	1780	1	YK26_CAEEL
4	77	17.1	482	1	R167_YEAST
5	75	16.6	1032	1	S242_HUMAN
6	73.5	16.3	494	1	SOX9_CHICK
7	73.5	16.3	817	1	NEB2_RAT
8	72.5	16.1	339	1	LSPI_HUMAN
9	72.5	16.1	670	1	SYN1_MOUSE
10	72.5	16.1	704	1	SYN1_RAT
11	72.5	16.1	705	1	SYN1_HUMAN
12	72.5	16.1	705	1	SYN1_BOVIN
13	72.5	16.1	719	1	DEND_RAT
14	72	16.0	310	1	Y497_MYCTU
15	72	16.0	2158	1	MY9B_HUMAN
16	72	16.0	3164	1	TEGU_HSV11
17	71.5	15.9	427	1	ODP2_BACST
18	71	15.7	861	1	PQ58_CAEEL
19	71	15.7	1443	1	SYJ2_HUMAN
20	70.5	15.6	2142	1	BAT2_HUMAN
21	70	15.5	943	1	NFC1_HUMAN
22	70	15.5	975	1	CUT1_CANFA
23	70	15.5	1274	1	ENAM_MOUSE
24	70	15.5	1298	1	ICP4_HSV11
25	69.5	15.4	415	1	SYN1_CANFA
26	69.5	15.4	700	1	MYBB_HUMAN
27	69.5	15.4	979	1	Y064_MYCTU
28	69.5	15.4	1319	1	MN1_HUMAN
29	69	15.3	1509	1	GSRI_HUMAN
30	68.5	15.2	359	1	Y40E_RHISN
31	68.5	15.2	509	1	SOX9_HUMAN
32	68.5	15.2	604	1	MAON_HUMAN
33	68.5	15.2	1238	1	BVGS_BORBR

34	68.5	15.2	1238	1	BVGS_BORPA	P40330 bordetella
35	68.5	15.2	1409	1	AEX3_CAEEL	O02626 caenorhabdi
36	68	15.1	628	1	V70K_TYINV	P10357 turnip yell
37	68	15.1	1026	1	STAU_DROME	P25159 drosophila
38	67.5	15.0	615	1	MUTL_ECOLI	P23367 escherichia
39	67	14.9	113	1	N12B_MEDSA	Q40339 medicago sa
40	67	14.9	296	1	PRP3_MOUSE	P05143 mus muscucu
41	67	14.9	309	1	HXA4_CHICK	P17277 gallus gall
42	67	14.9	421	1	FXJ1_HUMAN	Q92949 homo sapien
43	67	14.9	615	1	LM06_HUMAN	Q43900 homo sapien
44	67	14.9	622	1	3BP1_HUMAN	Q9Y313 homo sapien
45	67	14.9	1647	1	SN24_HUMAN	P51532 homo sapien

#### ALIGNMENTS

RESULT 1	TRLT_MACFA	STANDARD;	PRT;	430 AA.
ID	TRLT_MACFA	STANDARD;	PRT;	430 AA.
AC	Q2N092;			
DT	15-JUN-2002 (Rel. 41, Created)			
DT	15-JUN-2002 (Rel. 41, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Tumor necrosis factor receptor superfamily member 1:FRSF19L precursor			
DE	(Receptor expressed in lymphoid tissues).			
GN	TNFRSF19L OR RELT.			
OS	Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;			
OC	Cercopitheidae; Macaca.			
OX	NCBI_TaxID=9541;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Brain;			
RX	MEDLINE=21458551; PubMed=11574149;			
RA	Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirata M., Suto Y.,			
RA	Hirai M., Terao K., Suzuki Y., Sugano S., Hashimoto K., Kusuda J.;			
RT	"Assignment of 118 novel cDNAs of cynomolgus monkey brain to human			
RT	chromosomes."			
RL	Gene 27531-37(2001).			
CC	-1- FUNCTION: Mediates activation of NF-kappa-B (by similarity). May			
CC	play a role in T-cell activation.			
CC	-1- SUBUNIT: Associates with TRAF1 (By similarity).			
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein (Probable).			
CC	-1- SIMILARITY: CONTAINS 1 TNFR-CYS REPEAT.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
CC	EMBL: AB046039; BAB01621.1;			
DR	InterPro: IPR001368; TNFR_c6.			
DR	PROSITE: PS00652; TNFR_NGFR_1; FALSE_NEG.			
DR	PROSITE: PS00050; TNFR_NGFR_2; FALSE_NEG.			
DR	SMART: SM00208; TNFR; 1.			
KW	Receptor; Transmembrane; Glycoprotein; Signal.			
FT	SIGNAL 1 26			
FT	CHAIN 27 430			
FT	BY SIMILARITY.			
FT	TUMOR NECROSIS FACTOR RECEPTOR			
FT	SUPERFAMILY MEMBER TNFRSF19L.			
FT	EXTRACELLULAR (POTENTIAL).			
FT	POTENTIAL.			
FT	TRANSMEM			
FT	DOMAIN 163 183			
FT	DOMAIN 184 430			
FT	CYTOPLASMIC (POTENTIAL).			
FT	REPEAT 50 90			
FT	TNFR-CYS.			
FT	DISULFID 51 65			
FT	BY SIMILARITY.			
FT	DISULFID 71 90			
FT	BY SIMILARITY.			
FT	CARBOHYD 149 149			
FT	N-LINKED (GLCNAC. . .) (POTENTIAL).			
FT	SEQUENCE 430 AA; 45850 MW; BA8DE92593E1E859 CKC54;			

```
Query Match 18.3%; Score 82.5; DB 1; Length 430;
Best Local Similarity 28.2%; Pred. NO. 0.95;
Matches 22; Conservative 11; Mismatches 38; Indels 7; Gaps 3;

QY 7 PRVEGVP-VAVHKHALH--DGLRQVAGPGAAAHLPWP-----PPQLAASRRREAPPLSOR 59
Db 266 PAPPNPVPHICPHRHHLHTVOGLASLGGCCSCRSCKKWPVLLSPPEAVAATTSPSFLPN 325

QY 60 PRTQAGSPPTNEKLT 77
Db 326 PTRVPKAGAKAGQGEIT 343

RESULT 2
TRLT_HUMAN
ID TRLT_HUMAN STANDARD; PRT; 430 AA.
AC Q969Z4; Q96JUI; Q9BUX7;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Tumor necrosis factor receptor superfamily member TNFRSF19L precursor
DE (Receptor expressed in lymphoid tissues).
GN TNFRSF19L OR RELT.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF N-TERMINUS.
RC TISSUE=Lymphoma;
RA Sica G.L., Zhu G., Tamada K., Liu D., Ni J., Chen L.;
RA MEDLINE=21213541; PubMed=11313261;
RT "RELT, a new member of the tumor necrosis factor receptor superfamily,
RT is selectively expressed in hematopoietic tissues and activates
RT transcription factor NF-kappaB.";
RL Blood 97:2702-2707(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Retinoblastoma;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugeno S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi K., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwavanagi T.;
RT "NEO human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon, and Eye;
RA Strausberg R.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 121-430 FROM N.A.
RC TISSUE=Spleen;
RA Jikuya H., Takano J., Nomura N., Kikuno R., Nagase T., Ohara O.;
RT "The nucleotide sequence of a long cDNA clone isolated from human
RT spleen.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Mediates activation of NF-kappa-B. May play a role in T-
CC cell activation.
CC -!- SUBUNIT: Associates with TRAF1.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Probable).
CC -!- TISSUE SPECIFICITY: Highest levels are in spleen, lymph node,
CC thymus, peripheral blood leukocytes, bone marrow and fetal liver.
CC Very low levels in skeletal muscle, testis and colon. Not detected
CC in brain, kidney and pancreas.
CC -!- SIMILARITY: CONTAINS 1 TNFR-CYS REPEAT.
CC -!- CAUTION: Ref.4 sequence differs from that shown due to several
CC frameshifts.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF319553; AAK77356.1; -
DR EMBL; AK027899; BAB55441.1; -
DR EMBL; BC001812; AAH01812.1; -
DR EMBL; BC017279; AAH17279.1; -
DR EMBL; AK074128; BAB84954.1; ALT_FRAME.
DR Genew; HGNC:13764; TNFRSF19L.
DR PROSITE; PS00652; TNFR_NGFR_1; FALSE_NEG.
DR PROSITE; PS00652; TNFR_NGFR_2; FALSE_NEG.
KW Receptor; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 26
FT CHAIN 27 430 TUMOR NECROSIS FACTOR RECEPTOR
FT SUPERFAMILY MEMBER TNFRSF19L.
FT EXTRACELLULAR (POTENTIAL).
FT POTENTIAL.
FT CYTOPLASMIC (POTENTIAL).
FT TNFR-CYS.
FT BY SIMILARITY.
FT DISULFID 51 65
FT DISULFID 71 90
FT CARBOHYD 149 149 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 122 122 D -> S (IN REF. 4).
FT CONFLICT 187 187 K -> E (IN REF. 2).
FT CONFLICT 273 273 H -> R (IN REF. 2).
FT CONFLICT 379 380 DL -> TR (IN REF. 3; AAH01812).
FT SEQUENCE 430 AA; 46092 MW; 4A5AB9AE32D36101 CRC64;
Query Match 18.1%; Score 81.5; DB 1; Length 430;
Best Local Similarity 28.2%; Pred. NO. 1.2;
Matches 22; Conservative 11; Mismatches 38; Indels 7; Gaps 3;

QY 7 PRVEGVP-VAVHKHALH--DGLRQVAGPGAAAHLPWP-----PPQLAASRRREAPPLSOR 59
Db 266 PAPPNPVPHICPHRHHLHTVOGLASLGGCCSCRSCKKWPVLLSPPEAVAATTSPSFLPN 325

QY 60 PRTQAGSPPTNEKLT 77
Db 326 PTRVPKAGAKAGQGEIT 343

RESULT 3
YKZ6_CAEEL
ID YKZ6_CAEEL STANDARD; PRT; 1780 AA.
AC P34333; P34332;
DT 01-FEB-1994 (Rel. 28, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein C14B9.6 in chromosome III.
GN C14B9.6/C14B9.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Lathille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
RA Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston J., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Woldman P.;
```



16-OCT-2001 (Rel. 40, Last sequence update)  
16-OCT-2001 (Rel. 40, Last annotation update)  
Protein transport protein Sec24D (SEC24-related protein D).  
SEC24D OR KIAA0755.  
Homo sapiens (Human).  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
NCBI\_TaxID=9606;  
[1]  
SEQUENCE FROM N.A.  
TISSUE=Brain;  
MEDLINE=99087487; PubMed=9872452;  
Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,  
Tanaka A., Kotani H., Nomura N., Ohara O.;  
"Prediction of the coding sequences of unidentified human genes. XI.  
The complete sequences of 100 new cDNA clones from brain which code  
for large proteins in vitro.";  
DNA Res. 5:277-286(1998).  
[2]  
SEQUENCE FROM N.A.  
TISSUE=Pancreas;  
MEDLINE=99262157; PubMed=10329445;  
Tang B.-L., Kausalya J., Low D.Y.H., Lock M.L., Hong W.;  
"A family of mammalian proteins homologous to yeast Sec24p.";  
Biochem. Biophys. Res. Commun. 258:679-684(1999).  
- FUNCTION: COMPONENT OF THE COPII COAT, THAT COVERS ER-DERIVED  
VESICLES INVOLVED IN TRANSPORT FROM THE ENDOPLASMIC RETICULUM TO  
THE GOLGI APPARATUS. COPII ACTS IN THE CYTOPLASM TO PROMOTE THE  
TRANSPORT OF SECRETORY, PLASMA MEMBRANE, AND VACUOLAR PROTEINS  
FROM THE ENDOPLASMIC RETICULUM TO THE GOLGI COMPLEX.  
- SUBUNIT: COPII IS COMPOSED OF AT LEAST FIVE PROTEINS: THE SEC23/24  
COMPLEX, THE SEC13/31 COMPLEX AND SAR1.  
- SUBCELLULAR LOCATION: CYTOPLASMIC AND PERINUCLEAR (BY SIMILARITY).  
- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED, WITH HIGHER AMOUNTS IN  
PLACENTA, PANCREAS, HEART AND LIVER.  
- SIMILARITY: BELONGS TO THE SEC23/SEC24 FAMILY. SEC24 SUBFAMILY.  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
EMBL; AB018298; BAA34475.1; -  
EMBL; AF130464; AAD28756.2; -  
Genew; HGNC:10706; SEC24D.  
Transport; Protein transport; Golgi stack; Endoplasmic reticulum;  
Multigene family.  
DOMAIN 363  
DOMAIN 10 293  
SEQUENCE 1032 AA; 112999 MW; 1EAA80215EB979A7 CRC64;  
Query Match 16.6%; Score 75; DB 1; Length 1032;  
Best Local Similarity 35.0%; Pred. No. 11;  
Matches 21; Conservative 10; Mismatches 13; Indels 16; Gaps 4;  
QY 30 GPCAAAHLPW-----PPOLAASRRAPPSRPHRTQAGSPPTNEKLTNPQV 81  
DB 180 GPCASPLPLWYRDPDGLSGPPDPN---AOYQPPPL---PGQTLGAGYPPQ--QANSQPQM 231  
RESULT 6  
SOX9\_CHICK STANDARD; PRT; 494 AA.  
AC P48434;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Transcription factor SOX-9.  
GN SOX9.  
OS Gallus gallus (Chicken).

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauaria; Aves; Neognathae; Galliformes; Phasianinae;  
Gallus.  
NCBI\_TaxID=9031;  
[1]  
SEQUENCE FROM N.A.  
Uwanogho D., Rex M., Cartwright E.J., Pearl G., Scotting P.J.,  
Sharpe P.T.;  
Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
- SUBCELLULAR LOCATION: Nuclear (Potential).  
- SIMILARITY: CONTAINS 1 HMG BOX.  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
EMBL; U12533; AAB09663.1; -  
HSP; P48436; LSX9.  
InterPro: IPR000910; HMG\_12\_Box.  
Pfam: PF00505; HMG\_Box; 1.  
SMART: SM00398; HMG; 1.  
Transcription regulation; DNA-binding; Nuclear protein.  
FT DNA\_BIND 106 174 HMG\_BOX.  
FT DOMAIN 215 220 POLY-SER.  
FT DOMAIN 342 360 GLN/PRO-RICH.  
FT DOMAIN 394 401 POLY-GLN.  
FT SEQUENCE 494 AA; 54942 MW; AF41D483B5016349 CRC64;  
Query Match 16.3%; Score 73.5; DB 1; Length 494;  
Best Local Similarity 27.5%; Pred. No. 7.4;  
Matches 25; Conservative 12; Mismatches 37; Indels 17; Gaps 3;  
QY 11 GPVVA---VHKHALDGLRQVAGPGCAAAHL-----PRWPPPPQ--LAASRREA 53  
DB 307 GVPATHGQVITYSGTYGSSSSPAGAGHAWMAKQPPQPPQAPPAHQHTLPSTEREQ 366  
QY 54 PPLSRPHRTQAGSPPTNEKLTNPQVKEK 84  
DB 367 GPAQQRPHIKTEQLSPSHNSEQQQHPEQQQ 397  
RESULT 7  
NEB2\_RAT STANDARD; PRT; 817 AA.  
ID NEB2\_RAT  
AC O35274;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Neurabin-II (Neural tissue-specific F-actin binding protein II).  
DE (Protein phosphatase 1 regulatory subunit 9B) (Spinephillin) (p130)  
DE (Pp1bpl34).  
GN PP1R9B.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
NCBI\_TaxID=10116;  
[1]  
SEQUENCE FROM N.A.  
RX MEDLINE=97420791; PubMed=9275233;  
RT Allen P.B., Ouimet C.C., Greengard P.;  
"Spinophilin, a novel protein phosphatase 1 binding protein localized  
to dendritic spines.";  
Proc. Natl. Acad. Sci. U.S.A. 94:9956-9961(1997).  
[2]  
SEQUENCE FROM N.A., AND SEQUENCE OF 4-24; 164-182; 199-242; 449-480;  
RP 678-689 AND 776-791.  
RC TISSUE=Brain;  
RX MEDLINE=98123121; PubMed=9452470;  
RA Satoh A., Nakanishi H., Obaishi H., Wada M., Takahashi K., Satoh K.,







```

CC CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; SYNAPSIN IA (SHOWN HERE) AND
CC CC SYNAPSIN IB; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC CC -!- PTM: SUBSTRATE OF AT LEAST FOUR DIFFERENT PROTEIN KINASES. IT IS
CC CC PROBABLE THAT PHOSPHORYLATION PLAYS A ROLE IN THE REGULATION OF
CC CC SYNAPSIN I IN THE NERVE TERMINAL (BY SIMILARITY).
CC CC -!- SIMILARITY: BELONGS TO THE SYNAPSIN FAMILY.
CC CC -!- CAUTION: REF.2 SEQUENCE WAS INCORRECT AT MANY PLACES DUE TO
CC CC SEQUENCING ERRORS.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
DR DR EMBL; M27812; AAA42145.1; -
DR DR EMBL; M27924; AAA42148.1; -
DR DR EMBL; X04655; CAA28353.1; ALT_SEQ.
DR DR PIR; A25704; A25704.
DR DR PIR; A30411; A30411.
DR DR PIR; B30411; B30411.
DR DR HSSP; P17599; LAUX.
DR DR InterPro; IPR001359; Synapsin.
DR DR Pfam; PF02078; Synapsin; 1.
DR DR Pfam; PF02750; Synapsin_C; 1.
DR DR PRINTS; PR01368; SYNAPSIN.
DR DR PROSITE; PS00415; SYNAPSIN_1; 1.
DR DR PROSITE; PS00416; SYNAPSIN_2; 1.
DR DR Synapse; Phosphorylation; Nerve; Neurone; Repeat; Actin-binding;
KW Alternative splicing.
FT DOMAIN 1 28 A. (LINKER).
FT DOMAIN 29 112 B C (ACTIN-BINDING AND SYNAPTIC-VESICLE
FT DOMAIN 113 420 BINDING).
FT FT D (PRO-RICH LINKER).
FT FT E.
FT FT DOMAIN 421 655 PHOSPHORYLATION (BY PKA AND CAMK1).
FT FT DOMAIN 656 704 PHOSPHORYLATION (BY CAMK2).
FT FT MOD_RES 9 566 PHOSPHORYLATION (BY CAMK2).
FT FT MOD_RES 566 603 PHOSPHORYLATION (BY CAMK2).
FT FT MOD_RES 603 668 NKSQSLTNA -> KASPAQAOP (IN SYNAPSIN IB).
FT FT VARSPLIC 669 704 MISSING (IN SYNAPSIN IB).
FT VARSPLIC 669 704 MISSING (IN SYNAPSIN IB).
SQ SEQUENCE 704 AA; 73388 MW; 65799FEF7CFE18B5 CRC64;

Query Match 16.18; Score 72.5; DB 1; Length 704;
Best Local Similarity 35.8%; Pred. No. 13;
Matches 19; Conservative 7; Mismatches 18; Indels 9; Gaps 2;

Qy 26 QVAGPGAAAHLPWPQPQLA-ASRREAPPLSQPRTGAGSPETNEKLT 77
:| ||| | | : ||| :| ||| |||
Db 448 QQPAGPPAQRPQPGPGPPGPGPGPQPGPPLQQRP-----PPQGQOHL 492

RESULT 11
SYN1_HUMAN
ID SYN1_HUMAN STANDARD; PRT; 705 AA.
AC P17600; 075825;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Synapsin I (Brain protein 4.1).
GN SYN1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA MEDLINE-90243651; PubMed-2110562;
RA Suedhof T.C.;
RT "The structure of the human synapsin I gene and protein.";
```

```

RL J. Biol. Chem. 265:7849-7852(1990).
RN [2]
RP SEQUENCE OF 1-125 FROM N.A.
RX MEDLINE-90368667; PubMed-2118519;
RA Sauerwald A., Hoesch C., Oschwald R., Killmann M.W.;
RT "The 5'-flanking region of the synapsin I gene. A G-C-rich, TATA- and
RT CAAT-less, phylogenetically conserved sequence with cell
RT type-specific promoter function.";
RL J. Biol. Chem. 265:14932-14937(1990).
RN [3]
RP SEQUENCE OF 1-258 FROM N.A.
RA Grafham D.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: NEURONAL PHOSPHOPROTEIN THAT COATS SYNAPTIC VESICLES,
CC BINDS TO THE CYTOSKELETON, AND IS BELIEVED TO FUNCTION IN THE
CC REGULATION OF NEUROTRANSMITTER RELEASE.
CC -!- SUBCELLULAR LOCATION: SYNAPSE.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; SYNAPSIN IA (SHOWN HERE) AND
CC SYNAPSIN IB; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- PTM: SUBSTRATE OF AT LEAST FOUR DIFFERENT PROTEIN KINASES. IT IS
CC PROBABLE THAT PHOSPHORYLATION PLAYS A ROLE IN THE REGULATION OF
CC SYNAPSIN I IN THE NERVE TERMINAL.
CC -!- SIMILARITY: BELONGS TO THE SYNAPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC -----
DR DR EMBL; M58378; AAC41930.1; -
DR DR EMBL; M58321; AAC41930.1; JOINED.
DR DR EMBL; M58341; AAC41930.1; JOINED.
DR DR EMBL; M58351; AAC41930.1; JOINED.
DR DR EMBL; M58353; AAC41930.1; JOINED.
DR DR EMBL; M58359; AAC41930.1; JOINED.
DR DR EMBL; M58371; AAC41930.1; JOINED.
DR DR EMBL; M58372; AAC41930.1; JOINED.
DR DR EMBL; M58373; AAC41930.1; JOINED.
DR DR EMBL; M58374; AAC41930.1; JOINED.
DR DR EMBL; M58375; AAC41930.1; JOINED.
DR DR EMBL; M58376; AAC41930.1; JOINED.
DR DR EMBL; M58377; AAC41930.1; JOINED.
DR DR EMBL; M58378; AAC41930.1; JOINED.
DR DR EMBL; M58321; AAC41931.1; JOINED.
DR DR EMBL; M58341; AAC41931.1; JOINED.
DR DR EMBL; M58351; AAC41931.1; JOINED.
DR DR EMBL; M58353; AAC41931.1; JOINED.
DR DR EMBL; M58359; AAC41931.1; JOINED.
DR DR EMBL; M58371; AAC41931.1; JOINED.
DR DR EMBL; M58372; AAC41931.1; JOINED.
DR DR EMBL; M58373; AAC41931.1; JOINED.
DR DR EMBL; M58374; AAC41931.1; JOINED.
DR DR EMBL; M58375; AAC41931.1; JOINED.
DR DR EMBL; M58376; AAC41931.1; JOINED.
DR DR EMBL; M58377; AAC41931.1; JOINED.
DR DR EMBL; M55301; AAA60608.1; -
DR DR EMBL; AL009172; CAA15657.1; -
DR DR PIR; A35363; A35363.
DR DR PIR; B35363; B35363.
DR DR HSSP; P17599; LAUX.
DR DR Genew; HGNC:11494; SYN1.
DR DR MIM; 313440; -
DR DR InterPro; IPR001359; Synapsin.
DR DR Pfam; PF02078; Synapsin; 1.
DR DR Pfam; PF02750; Synapsin_C; 1.
DR DR PRINTS; PR01368; SYNAPSIN.
DR DR PROSITE; PS00415; SYNAPSIN_1; 1.
DR DR PROSITE; PS00416; SYNAPSIN_2; 1.
KW Synapse; Phosphorylation; Neurone; Repeat; Actin-binding;
KW Alternative splicing.
```

```

FT DOMAIN 1 28
FT DOMAIN 29 112
FT DOMAIN 113 420
FT DOMAIN 421 655
FT DOMAIN 656 705
FT MOD_RES 9 9
FT MOD_RES 568 568
FT MOD_RES 605 605
FT MOD_RES 661 669
FT VARSPLIC 670 705
FT VARSPLIC 138 138
SQ SEQUENCE 705 AA; 73954 MW; 487831123FF6882F CRC64;

Query Match 16.1%; Score 72.5; DB 1; Length 705;
Best Local Similarity 35.8%; Pred. No. 13;
Matches 19; Conservative 7; Mismatches 18; Indels 9; Gaps 2;

QY 26 QVAGGAAAHLPWPQOLA-ASRREAPLSQSRHRTQAGSPETNEKLT 77
Db 450 QQAGPPAQRPPQGGPPQPGPQRCQPPQLQRRP-----PPQGGQHLS 494

RESULT 12
SYN1_BOVIN
ID SYN1_BOVIN STANDARD; PRT; 706 AA.
AC P17599;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Synapsin I.
GN SYN1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=89388265; PubMed=2506642;
RA Suedhof T.C., Czernik A.J., Kao H.-T., Takei K., Johnston P.A.,
RA Horluchi A., Kanazir S.D., Wagner M.A., Perin M.S., de Camilli P.,
RA Greengard P.;
RT "Synapsins: mosaics of shared and individual domains in a family of
RT synaptic vesicle phosphoproteins.";
RL Science 245:1474-1480(1989).
RN [2]
RP PHOSPHORYLATION SITE BY PROLINE-DIRECTED KINASE.
RX MEDLINE=90216728; PubMed=2108963;
RA Hall F.L., Mitchell J.P., Vulliamt P.R.;
RT "Phosphorylation of synapsin I at a novel site by proline-directed
RT protein kinase.";
RL J. Biol. Chem. 265:6944-6948(1990).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS) OF 112-417.
RX MEDLINE=98130589; PubMed=9463376;
RA Esser L., Wang C.R., Hosaka M., Smagula C.S., Suedhof T.C.,
RA Deisenhofer J.;
RT "Synapsin I is structurally similar to ATP-utilizing enzymes.";
RL EMBO J. 17:977-984(1998).
CC -!- FUNCTION: NEURONAL PHOSPHOPROTEIN THAT COATS SYNAPTIC VESICLES,
CC BINDS TO THE CYTOSKELETON, AND IS BELIEVED TO FUNCTION IN THE
CC REGULATION OF NEUROTRANSMITTER RELEASE.
CC -!- SUBCELLULAR LOCATION: SYNAPSE.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; SYNAPSIN IA (SHOWN HERE) AND
CC SYNAPSIN IB; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- PTM: SUBSTRATE OF AT LEAST FOUR DIFFERENT PROTEIN KINASES. IT IS
CC PROBABLE THAT PHOSPHORYLATION PLAYS A ROLE IN THE REGULATION OF
CC SYNAPSIN I IN THE NERVE TERMINAL.
CC -!- SIMILARITY: BELONGS TO THE SYNAPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M27810; AAA30761.1; -.
CC EMBL; M27811; AAA30762.1; -.
CC PIR; A35758; A35758.
CC PIR; E30411; E30411.
CC PIR; F30411; F30411.
CC PDB; 1AUV; 18-MAR-98.
CC PDB; 1AUX; 18-MAR-98.
CC GlycoSuiteDB; P17599; -.
CC InterPro; IPR001359; Synapsin.
CC Pfam; PF02078; Synapsin; 1.
CC Pfam; PF02750; Synapsin_C; 1.
CC PRINTS; PR01368; SYNAPSIN.
CC PROSITE; PS00415; SYNAPSIN_1; 1.
CC PROSITE; PS00416; SYNAPSIN_2; 1.
CC Synapse; Phosphorylation; Neurone; Repeat; Actin-binding;
KW Alternative splicing; 3D-structure.
FT DOMAIN 1 28
FT DOMAIN 29 112
FT DOMAIN 113 420
FT DOMAIN 421 656
FT DOMAIN 657 706
FT MOD_RES 9 9
FT MOD_RES 551 551
FT MOD_RES 568 568
FT MOD_RES 605 605
FT MOD_RES 662 670
FT VARSPLIC 671 706
SQ SEQUENCE 706 AA; 74518 MW; 89373750BF014340 CRC64;

Query Match 16.1%; Score 72.5; DB 1; Length 706;
Best Local Similarity 26.4%; Pred. No. 13;
Matches 28; Conservative 9; Mismatches 28; Indels 41; Gaps 5;

QY 5 APRVEGVVAVHKHALHDLRQVAGP-----GAAAHLPWPQOLA----- 46
Db 573 APPKAGVP-----PGQQRQGPQKPPGPGAGTQQAQAGPMPRTGPTTQOQRP 623
QY 47 ----AASRREAPLSQSRPHR-----TOGAGSP-----ETNEKLTN 78
Db 624 SGPGPAGRTKPOLAQKPSQDVPPPPATAAAGGPPHQLNKQSOLTN 669

RESULT 13
DEND_RAT
ID DEND_RAT STANDARD; PRT; 719 AA.
AC P50817;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Dendrin (Fragment).
GN DDN
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=97073443; PubMed=8915891;
RA Neuner-Jehle M., Denizot J.P., Borbely A.A., Mallet J.;
RT "Characterization and sleep deprivation-induced expression modulation
RT of dendrin, a novel dendritic protein in rat brain neurons.";
RL J. Neurosci. Res. 46:138-151(1996).
CC -!- MISCELLANEOUS: MODULATED BY SLEEP DEPRIVATION.
CC -----

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: X96589; CRA65407.1; -  
CC FT NON\_TER 1  
CC SQ SEQUENCE 719 AA; 77208 MW; 89E2ED094514EAFB CRC64;

Query Match 16.1%; Score 72.5; DB 1; Length 719;  
Best Local Similarity 33.9%; Pred. No. 13;  
Matches 21; Conservative 5; Mismatches 17; Indels 19; Gaps 3;

QY 30 GPGAAAHNL-----PRW--PPQLAASRREAPLSQRPHTQAGSPPTNEKL 76  
DB 211 GPPSYEAHLLLRGAAGTAPRRWRPPPV-----APPSYEGPHTLTGKRGPELSR 264  
QY 77 TN 78  
DB 265 TS 266

RESULT 14  
Y497\_MYCTU STANDARD; PRT; 310 AA.  
AC Q11162;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update).  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein RV0497.  
GN RV0497 OR MT0517 OR MTCY20G9.23.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=11773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37Rv;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Gordon S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Cole S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence.";  
RL Nature 393:537-544(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CDC 1551 / Oshkosh;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
RA Bishai W.;  
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains.";  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein (potential).  
CC -1- SIMILARITY: TO M.LEPRA MC2433.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: X96589; CRA65407.1; -  
CC FT NON\_TER 1  
CC SQ SEQUENCE 719 AA; 77208 MW; 89E2ED094514EAFB CRC64;

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: 277162; CAB00923.1; -  
CC DR EMBL; AE006952; AAK44740.1; -  
CC TIGR: MT0517; -  
CC DR Tuberculin; RV0497; -  
CC KW Hypothetical protein; Transmembrane; Complete proteome.  
CC FT TRANSMEM 231 251 POTENTIAL.  
CC FT TRANSMEM 257 277 POTENTIAL.  
CC FT TRANSMEM 286 306 POTENTIAL.  
CC FT DOMAIN 33 39 POLY-ARG.  
CC FT DOMAIN 197 202 POLY-ALA.  
CC SQ SEQUENCE 310 AA; 33092 MW; 4954027F694DF5C2 CRC64;

Query Match 16.0%; Score 72; DB 1; Length 310;  
Best Local Similarity 30.4%; Pred. No. 6.4;  
Matches 28; Conservative 9; Mismatches 29; Indels 26; Gaps 6;  
QY 4 DAAPRVEGYPVAVHKKHALHDGLRQVAGCGAAAHNL-----PRWP--PPQLAASRREA-PPL 56  
DB 84 EAAQSPAEPA-----EQVAEPTRTVYWSQPEPRWPKSPQ---DRRESGPPL 130  
QY 57 SORP-----HRTQAGSPPTNEKLTPQVK 82  
DB 131 SEYPRPLRTHSDRAPAGPPSGAEHMSDPVE 162

RESULT 15  
MY9B\_HUMAN STANDARD; PRT; 2158 AA.  
ID MY9B\_HUMAN  
AC Q13459; Q9UHN0; Q75314; Q9NUJ2;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Myosin IxB (unconventional myosin-9b).  
GN MYO9B OR MYR5.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver, and Small intestine;  
RX MEDLINE=97063843; PubMed=8907710;  
RA Wirth J.A., Jensen K.A., Post P.L., Bement W.M., Moosker M.S.;  
RT "Human myosin-IxB, an unconventional myosin with a chimerin-like  
RT rho/rac GTPase-activating protein domain in its tail.";  
RL J. Cell Sci. 109:653-661(1996).  
RN [2]  
RP PARTIAL SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
RX MEDLINE=20047919; PubMed=10580159;  
RA Grewal P.K., Jones A.-M., Maconochie M., Lemmers R.J.F., Frants R.R.,  
RA Hewitt J.E.;  
RT "Cloning of the murine unconventional myosin gene Myo9b and  
RT identification of alternative splicing.";  
RL Gene 240:389-398(1999).  
RN [3]  
RP SEQUENCE OF 1940-2158 FROM N.A. (LONG ISOFORM).  
RX MEDLINE=98158729; PubMed=9490638;  
RA Post P.L., Bokoch G.M., Moosker M.S.;  
RT "Human myosin-IxB is a mechanoschemically active molecule and a GAP for  
RT rho.";  
RL J. Cell Sci. 111:941-950(1998).  
RN [4]  
RP SEQUENCE OF 1828-2023 FROM N.A. (SHORT ISOFORM).  
RC TISSUE=Placenta;  
RA Ito T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Ishibashi T., Fujimori K.,  
RA Tanai H., Kinata M., Watanabe M., Hiraoka S., Ishii S., Kawai Y.,  
RA Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K.,  
RA Masuho Y., Kanehori K.;  
RT "NEDO human cDNA sequencing project.";  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: MYOSINS ARE ACTIN-BASED MOTOR MOLECULES WITH ATPASE  
CC ACTIVITY. UNCONVENTIONAL MYOSINS SERVE IN INTRACELLULAR MOVEMENTS.  
CC MAY BE INVOLVED IN THE REMODELING OF THE ACTIN CYTOSKELETON. BINDS  
CC ACTIN WITH HIGH AFFINITY BOTH IN THE ABSENCE AND PRESENCE OF ATP  
CC AND ITS MECHANOCHEMICAL ACTIVITY IS INHIBITED BY CALCIUM IONS.  
CC ALSO ACTS AS A GPINSE ACTIVATING PROTEIN ON RHO.  
CC -!- SUBCELLULAR LOCATION: IN UNDIFFERENTIATED CELLS COLOCALIZES WITH  
CC F-ACTIN IN THE CELL PERIPHERY WHILE IN DIFFERENTIATED CELLS ITS  
CC LOCALIZATION IS CYTOPLASMIC WITH THE HIGHEST LEVELS IN THE  
CC PERINUCLEAR REGION.  
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A  
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -!- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN PERIPHERAL BLOOD  
CC LEUCOCYTES AND AT LOWER LEVELS, IN THYMUS, SPLEEN, TESTIS,  
CC PROSTATE, OVARY, BRAIN, SMALL INTESTINE, AND LUNG.  
CC -!- SIMILARITY: CONTAINS 1 ZINC-DEPENDENT PHORBOL-ESTER AND DAG  
CC BINDING DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 RHO-GAP DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.  
CC -!- SIMILARITY: CONTAINS 4 IQ DOMAINS.  
CC -!- CAUTION: THE C-TERMINAL SEQUENCE FROM AA 1917 ONWARDS FROM REF.1  
CC WAS PROBABLY A CHIMERA.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
-----

DR EMBL; U42391; AAC50402.1; ALT\_SEQ.  
DR EMBL; AF143684; AAF00119.1; -  
DR EMBL; AF020267; AAC26597.1; -  
DR EMBL; AK002201; BAA92132.1; ALT\_INIT.  
DR HSSP; P08799; 1MND.  
DR Genew; HGNC:7609; MYO9B.  
DR MIM; 602129; -  
DR InterPro; IPR002219; DAG\_PE-bind.  
DR InterPro; IPR000048; IQ\_region.  
DR InterPro; IPR000159; RA\_domain.  
DR InterPro; IPR000198; RhoGAP.  
DR InterPro; IPR001609; myosin\_head.  
DR Pfam; PF00063; myosin\_head; 2.  
DR Pfam; PF00130; DAG\_PE-bind; 1.  
DR Pfam; PF00612; IQ; 4.  
DR Pfam; PF00620; RhoGAP; 1.  
DR Pfam; PF00788; RA; 1.  
DR PRINTS; PR00193; MYOSINHEAVY.  
DR ProDom; PD000355; myosin\_head; 2.  
DR SMART; SM00109; C1; 1.  
DR SMART; SM00015; IQ; 3.  
DR SMART; SM00242; MYSC; 1.  
DR SMART; SM00314; RA; 1.  
DR SMART; SM00324; RhoGAP; 1.  
DR PROSITE; PS00479; DAG\_PE\_BIND\_DOM\_1; 1.  
DR PROSITE; PS50081; DAG\_PE\_BIND\_DOM\_2; 1.  
DR PROSITE; PS50096; IQ; 3.  
KW Myosin; Repeat; ATP-binding; Calmodulin-binding; Actin-binding;  
KW Cytoskeleton; Coiled coil; GTPase activation; Phorbol-ester binding;  
KW Zinc; Alternative splicing.  
FT DOMAIN 1 939 HEAD OR MOTOR DOMAIN.  
FT DOMAIN 940 1044 NECK OR REGULATORY DOMAIN.  
FT DOMAIN 1045 2158 TAIL.  
FT DOMAIN 1046 1071 COILED COIL (POTENTIAL).  
FT DOMAIN 1880 1901 COILED COIL (POTENTIAL).  
FT DOMAIN 1959 1989 COILED COIL (POTENTIAL).  
FT DOMAIN 844 855 ACTIN-BINDING.  
FT DOMAIN 957 977 IQ 1.  
FT DOMAIN 979 1000 IQ 2.  
FT DOMAIN 1001 1023 IQ 3.  
FT DOMAIN 1024 1053 IQ 4.  
FT DOMAIN 1633 1681 PHORBOL-ESTER AND DAG BINDING.

FT DOMAIN 1713 1862 RHO-GAP.  
FT NP\_BIND 239 246 ATP (POTENTIAL).  
FT VARSPIC 2022 2023 PP -> QY (IN SHORT ISOFORM).  
FT VARSPIC 2024 2158 MISSING (IN SHORT ISOFORM).  
FT CONFLICT 1937 1939 QVP -> KT (IN REF. 4).  
FT CONFLICT 1947 1947 L -> P (IN REF. 4).  
FT CONFLICT 2040 2045 TVAAP -> PMPPLH (IN REF. 3).  
FT CONFLICT 2049 2049 P -> L (IN REF. 3).  
FT CONFLICT 2067 2067 P -> S (IN REF. 3).  
FT CONFLICT 2157 2158 NG -> MAESHS (IN REF. 3).  
SQ SEQUENCE 2158 AA; 243556 MW; 4978F1D770F56D28 CRC64;  
  
Query Match 16.0%; Score 72; DB 1; Length 2158;  
Best Local Similarity 29.2%; Pred. No. 43;  
Matches 21; Conservative 8; Mismatches 19; Indels 24; Gaps 4;  
  
QY 31 PGAAAAHLPRWPPQLAAS---RREAPPLSORPHR-----TQGA-----G 67  
| :||||| : || || :|| :  
Db 2074 PGGLPSHLPRWAPCARAAAPVRRRPP-ARRPDQHSVITPGADLPVQCALEPLEEDG 2132  
  
QY 68 SPPETNEKLTNP 79  
|| : :||  
Db 2133 QPPGAKRRYSDP 2144

Search completed: November 26, 2002, 12:02:21  
Job time : 40.0968 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 26, 2002, 12:00:06 ; Search time 29.8064 seconds  
(without alignments)  
82.919 Million cell updates/sec

Title: US-09-674-593-2

Perfect score: 451

Sequence: 1 MDDDAAPRVGVPAVHRHA.....GAGSPETNEKLTNPQVREK 84

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2.6/ptodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2.6/ptodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2.6/ptodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2.6/ptodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2.6/ptodata/1/1aa/PTUS\_COMB.pep:\*  
6: /cgn2.6/ptodata/1/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	77.5	17.2	1099	US-09-442-100-2	Sequence 2, Appl
2	72.5	16.1	696	US-08-906-865-4	Sequence 4, Appl
3	72.5	16.1	696	US-09-129-668-4	Sequence 4, Appl
4	70	15.5	1298	US-08-690-473-2	Sequence 2, Appl
5	70	15.5	1298	US-09-259-821A-2	Sequence 2, Appl
6	70	15.5	1298	US-08-843-659-2	Sequence 2, Appl
7	68.5	15.2	509	US-08-860-635A-21	Sequence 21, Appl
8	68.5	15.2	509	US-09-281-476-21	Sequence 21, Appl
9	68	15.1	354	US-09-608-917A-2	Sequence 2, Appl
10	67.5	15.0	615	US-08-676-444-44	Sequence 44, Appl
11	67	14.9	543	US-09-535-008-63	Sequence 63, Appl
12	67	14.9	577	US-09-535-008-61	Sequence 61, Appl
13	67	14.9	1646	US-09-535-008-67	Sequence 67, Appl
14	67	14.9	1647	US-09-535-008-2	Sequence 2, Appl
15	67	14.9	1649	US-09-535-008-75	Sequence 75, Appl
16	67	14.9	1650	US-09-535-008-71	Sequence 71, Appl
17	67	14.9	1678	US-09-535-008-69	Sequence 69, Appl
18	67	14.9	1679	US-09-535-008-65	Sequence 65, Appl
19	67	14.9	1681	US-09-535-008-77	Sequence 77, Appl
20	67	14.9	1682	US-09-535-008-73	Sequence 73, Appl
21	66.5	14.7	625	US-09-347-801-18	Sequence 18, Appl
22	66	14.6	516	US-09-291-170A-1	Sequence 1, Appl
23	66	14.6	516	US-09-724-884-1	Sequence 1, Appl
24	65.5	14.5	404	US-09-232-468A-8	Sequence 8, Appl
25	65.5	14.5	3025	US-09-232-468A-8	Patent No. 5223423
26	65	14.4	834	US-08-491-357-2	Sequence 2, Appl
27	65	14.4	834	US-08-968-633-2	Sequence 2, Appl

28 65 14.4 834 3 US-09-196-466-2 Sequence 2, Appl  
29 65 14.4 834 5 PCT-US96-10823-2 Sequence 2, Appl  
30 65 14.4 905 2 US-08-574-959A-9 Sequence 9, Appl  
31 65 14.4 905 4 US-09-357-014-9 Sequence 9, Appl  
32 65 14.4 1135 2 US-08-574-959A-7 Sequence 7, Appl  
33 65 14.4 1135 4 US-09-357-014-7 Sequence 7, Appl  
34 65 14.4 2414 1 PCT-US95-04682-2 Sequence 2, Appl  
35 65 14.4 2414 5 US-09-413-574-2 Sequence 2, Appl  
36 65 14.4 2860 2 US-08-826-267-2 Sequence 2, Appl  
37 64.5 14.3 865 4 US-09-281-766-19 Sequence 19, Appl  
38 64 14.2 265 4 US-09-199-637A-369 Sequence 369, Appl  
39 64 14.2 340 4 US-09-360-779-2 Sequence 2, Appl  
40 64 14.2 340 4 US-09-435-335-2 Sequence 2, Appl  
41 64 14.2 405 4 US-09-413-574-2 Sequence 2, Appl  
42 64 14.2 591 1 US-08-484-840-2 Sequence 2, Appl  
43 64 14.2 591 1 US-08-483-094-2 Sequence 2, Appl  
44 63.5 14.1 2293 4 US-09-368-590-2 Sequence 2, Appl  
45 63 14.0 432 2 US-08-937-466-4 Sequence 4, Appl

#### ALIGNMENTS

RESULT 1  
US-09-442-100-2  
; Sequence 2, Application US/09442100  
; Patent No. 6359193  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Tian  
; APPLICANT: Tao, Wufan  
; APPLICANT: Wang, Weiwei  
; APPLICANT: Zhang, Sheng  
; APPLICANT: Yu, Wan  
; TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF LATS  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/442,100  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/411,111  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mastro, S. Leslie  
; REGISTRATION NUMBER: 18,872  
; REFERENCE/DOCKET NUMBER: 6523-003  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-9741/8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1099 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; US-09-442-100-2  
Query Match 17.2%; Score 77.5; DB 4; Length 1099;  
Best Local Similarity 27.8%; Pred. No. 1.5;



```

Db      552  DDDAGGRRQAAGRVAVECLACACRGILEAEGFDGLAAV-----PGLAGAPASPP-- 602
QY      58   QRPRTQGAGSPP 70
Db      603  -RPEGPAGPASPP 614

```







```
; APPLICANT: Teng, David H.-F.
; TITLE OF INVENTION: BRG1 IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE
; FILE REFERENCE: 2318-259
; CURRENT APPLICATION NUMBER: US/09/535,008
; CURRENT FILING DATE: 2000-03-23
; EARLIER APPLICATION NUMBER: U.S. 60/125,806
; EARLIER FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 75
; LENGTH: 1649
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-535-008-75

Query Match      14.9%; Score 67; DB 4; Length 1649;
Best Local Similarity 27.4%; Pred. No. 37;
Matches 20; Conservative 6; Mismatches 31; Indels 16; Gaps 2;

QY 11 GVPVAVHKHALHDGLRQVAGPCAAAHLPWWPPOLAAS-----RREAPPLS 57
DB 193 GQPLPDHLQMAVQGRKMPG---MQQQMTLPPPSVATGPGPGPGPGPPNY 249
QY 58 QRPHTQAGAGSPP 70
DB 250 SRPHGCGGNMPP 262

Search completed: November 26, 2002, 19:43:18
Job time : 33.8064 secs
```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 26, 2002, 12:01:36 : Search time 44.2581 Seconds  
(without alignments)  
30.223 Million cell updates/sec

Title: US-09-674-593-2

Perfect score: 451

Sequence: 1 MDDDAAPRVGVPAVHKHA.....GAGSPETNEKLTNPQVKEK 84

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 102317 seqs, 15924203 residues

Total number of hits satisfying chosen parameters: 102317

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_AA.\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	81.5	18.1	408	10	US-09-057-951-4
2	81.5	18.1	408	12	US-10-105-150-4
3	81.5	18.1	430	10	US-09-057-951-2
4	81.5	18.1	430	10	US-09-836-607-2
5	81.5	18.1	430	12	US-10-105-150-2
6	72.5	16.1	193	10	US-09-864-761-37748
7	69.5	15.4	538	12	US-10-023-529-43
8	69.5	15.4	538	12	US-10-023-523-43
9	68.5	15.2	509	10	US-09-910-087-21
10	68	15.1	684	10	US-09-823-240-9
11	67	14.9	721	9	US-09-764-868-731
12	67	14.9	920	10	US-09-983-531A-4
13	66.5	14.7	625	10	US-09-854-731-18
14	66.5	14.7	846	10	US-09-858-664A-3
15	66.5	14.7	4019	10	US-09-738-973-425
16	65.5	14.5	531	10	US-09-925-300-1444
17	65	14.4	494	10	US-09-833-790-234
18	64.5	14.3	190	10	US-09-864-761-36101
19	64.5	14.3	419	10	US-09-814-777A-36

20	64.5	14.3	865	9	US-09-957-995A-19
21	64	14.2	340	10	US-09-850-799-2
22	64	14.2	405	10	US-09-805-550-2
23	64	14.2	430	12	US-10-062-254-252
24	64	14.2	888	12	US-10-036-342-35
25	63.5	14.1	104	10	US-09-864-761-33381
26	63.5	14.1	339	10	US-09-745-763-146
27	63.5	14.1	339	12	US-10-036-342-80
28	63.5	14.1	339	12	US-10-052-586-504
29	63.5	14.1	503	12	US-10-062-254-240
30	63	14.0	184	10	US-09-764-864-827
31	63	14.0	302	10	US-09-764-864-1331
32	63	14.0	520	10	US-09-726-643-102
33	63	14.0	526	10	US-09-925-300-1473
34	63	14.0	591	10	US-09-962-290-4
35	62.5	13.9	280	10	US-09-823-038A-18
36	62	13.7	507	10	US-09-910-087-19
37	62	13.7	559	10	US-09-877-633-12
38	62	13.7	596	12	US-10-006-867-100
39	62	13.7	596	12	US-10-052-586-310
40	61.5	13.6	283	12	US-10-052-586-194
41	61.5	13.6	321	9	US-09-945-182-26
42	61.5	13.6	480	10	US-09-790-264-32
43	61.5	13.6	756	9	US-09-992-598-62
44	61.5	13.6	756	10	US-09-989-722-62
45	61.5	13.6	756	10	US-09-989-723-62

ALIGNMENTS

RESULT 1  
US-09-057-951-4  
; Sequence 4, Application US/09057951  
; Patent No. US2002025551A1  
; GENERAL INFORMATION:  
; APPLICANT: Holtzman, Douglas  
; TITLE OF INVENTION: NOVEL MOLECULES OF THE T129-RELATED  
; TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/057,951  
; FILING DATE: 09-APR-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meiklejohn, Ph.D., Anita L.  
; REGISTRATION NUMBER: 35,283  
; REFERENCE/DOCKET NUMBER: 09404/046001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 408 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-057-951-4  
Query Match 18.1%; Score 81.5; DB 10: Length 408;  
Best Local Similarity 28.2%; Pred. No. 0.36;



Query Match	18.1%	Score 81.5;	DB 10;	Length 430;
Best Local Similarity	28.2%	Pred. No. 0.38;		
Matches	22;	Conservative	11;	Mismatches 38;
				Indels 7;
				Gaps

```

US-09-864-761-37748
; Sequence 37748, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 4917
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 37748
; LENGTH: 193
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC002467.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN HEPL100, SIGNAL = 2.1
; OTHER INFORMATION: EST_HUMAN HIT: AA831893.1, EVALUATE 7.00e-07
; US-09-864-761-37748

```



```
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 21:
US-09-010-087-21

Query Match      15.2%; Score 68.5; DB 10; Length 509;
Best Local Similarity 33.3%; Pred. No. 9.7;
Matches 24; Conservative 7; Mismatches 22; Indels 19; Gaps 6;

QY 11 GVPVAVHKHALHD---GLRVAGVGGAAAHLPW-----PPQLAASRRAPPPLSQS 59
Db 305 GVP-ATHGQVITYGSGISSTAATPASAGHV--WMSKQAPPVPPQ---OPQAPPAPQA 358
QY 60 PHTQAGAGSPE 71
Db 359 PPQPQAA--PPQ 368

RESULT 10
US-09-823-240-9
; Sequence 9, Application US/09823240
; Patent No. US20020048813A1
; GENERAL INFORMATION:
; APPLICANT: Frank B. Gertler
; APPLICANT: James E. Bear
; APPLICANT: Jurgen Wehland
; APPLICANT: Joseph Loureiro
; TITLE OF INVENTION: Methods and Products for Regulating Cell
; FILE REFERENCE: Motility
; CURRENT APPLICATION NUMBER: M0656/7064 (HCL)
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: US/09/823,240
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/194,564
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 684
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-823-240-9

Query Match      15.1%; Score 68; DB 10; Length 684;
Best Local Similarity 36.6%; Pred. No. 15;
Matches 15; Conservative 1; Mismatches 25; Indels 0; Gaps 0;

QY 30 GPGAAAHLPWPPPPQLAASRRAPPPLSQRPHRTQAGAGSPP 70
Db 400 GPGGPPAPAPPPPPPSFGGAAGGPPPPAPPQMFENGAPPPP 440

RESULT 11
US-09-764-868-731
; Sequence 731, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 731
; LENGTH: 721
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (123)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 21:
US-09-010-087-21

Query Match      14.9%; Score 67; DB 9; Length 721;
Best Local Similarity 34.7%; Pred. No. 20;
Matches 25; Conservative 4; Mismatches 29; Indels 14; Gaps 4;

QY 12 VPAVAVHKHALHDGLRQV-AGPGAAAHLPWPPPPQLAASR-----REAPP-LSORP 60
Db 461 VPAKSRERLANGLHPVPMGPSGA---LPSPDAPCLPVKRGSPASPTSPSPCPALAPRP 517
QY 61 HRTQAGAGSPPET 72
Db 518 LSGQAGGSPST 529

RESULT 12
US-09-983-531A-4
; Sequence 4, Application US/09983531A
; Patent No. US20020147132A1
; GENERAL INFORMATION:
; APPLICANT: Fujisawa, Atsuko
; APPLICANT: Yamakawa, Toru
; APPLICANT: Shirakawa, Kamon
; APPLICANT: Chitose, Orii
; APPLICANT: Ogawa, Naoki
; TITLE OF INVENTION: Meltrins
; FILE REFERENCE: 11-22-99 sequence submission
; CURRENT APPLICATION NUMBER: US/09/983,531A
; CURRENT FILING DATE: 2001-10-24
; PRIOR FILING DATE: 2001-10-24
; PRIOR APPLICATION NUMBER: JP 8-61756
; PRIOR FILING DATE: 1996-02-23
; PRIOR APPLICATION NUMBER: PCT/JP96/03017
; PRIOR FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 920
; TYPE: PRT
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Clone:
; OTHER INFORMATION: JM109(pBSmel-beta), mouse beta meltrin
; NAME/KEY: Unsure
; LOCATION: 574
; OTHER INFORMATION: Amino acid "Xaa" is unknown
US-09-983-531A-4

Query Match      14.9%; Score 67; DB 10; Length 920;
Best Local Similarity 36.6%; Pred. No. 27;
Matches 15; Conservative 5; Mismatches 19; Indels 2; Gaps 1;

QY 36 AHLPRWPPPPQLAASRRAPPPLSQRPHRTQAGAGSPPETNEKL 76
Db 789 SHPLRPPPPDYL--RVESPAPLPSAHLNRAAGSSPEAGARI 827

RESULT 13
US-09-854-731-18
; Sequence 18, Application US/09854731
; Patent No. US20020120949A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Lee, Jian Ming
; TITLE OF INVENTION: Plant Protein Kinases
; FILE REFERENCE: BB-1171
; CURRENT APPLICATION NUMBER: US/09/854,731
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: 60/092,438
; PRIOR FILING DATE: July 10, 1998
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 18
; LENGTH: 625
```

; TYPE: PRT  
; ORGANISM: Zea mays  
US-09-854-731-18

Query Match 14.7%; Score 66.5; DB 10; Length 625;  
Best Local Similarity 28.1%; Pred. No. 20;  
Matches 27; Conservative 2; Mismatches 26; Indels 41; Gaps 5;  
QY 5 AARPVAVHKHALHDLGVAGPGAAAHLPWWPPQLAAS-----49  
DB 42 ATPRHKSGSTTPVHH-----QAATPGAAA-----WPSYPAGGASPLPAGVSPSPA 89  
QY 50 -----RREAPPLSQRP-----RTQGAGSPPE 71  
DB 90 RSTPRFRKRPFPSPAKHKATLAKRLGGKPKPE 125

RESULT 14

US-09-858-664A-3  
; Sequence 3, Application US/09858664A  
; Patent No. US20020072491A1  
; GENERAL INFORMATION:  
; APPLICANT: WEI, Ming-Hui, et al.  
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
; FILE REFERENCE: CLO00927-CIP  
; CURRENT APPLICATION NUMBER: US/09/858,664A  
; CURRENT FILING DATE: 2001-05-17  
; PRIOR APPLICATION NUMBER: 09/711,134  
; PRIOR FILING DATE: 2000-11-11  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 846  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-858-664A-3

Query Match 14.7%; Score 66.5; DB 10; Length 846;  
Best Local Similarity 37.7%; Pred. No. 28;  
Matches 20; Conservative 0; Mismatches 32; Indels 1; Gaps 1;

QY 30 GPGAAAHLPWWPPQLAASRREA-PPLSQRPRTQGAGSPPETNEKLTNPQV 81  
DB 776 GFSQGAAPPSEPKPHAAVFARVASPPPGCAPEKRVPSAGGPPVLAEKARVPTV 828

RESULT 15

US-09-738-973-425  
; Sequence 425, Application US/09738973  
; Patent No. US20020110563A1  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Fling, Steven P.  
; APPLICANT: Mohamath, Raodoh  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Secrist, Heather  
; APPLICANT: Indrias, Carol Yoseph  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Elliot, Mark  
; APPLICANT: Mannion, Jane  
; APPLICANT: Kalos, Michael D.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.475C9  
; CURRENT APPLICATION NUMBER: US/09/738,973  
; CURRENT FILING DATE: 2000-12-14  
; NUMBER OF SEQ ID NOS: 587  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 425

; LENGTH: 4019  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-738-973-425

Query Match 14.7%; Score 66.5; DB 10; Length 4019;  
Best Local Similarity 31.1%; Pred. No. 1.7e+02;  
Matches 19; Conservative 5; Mismatches 18; Indels 19; Gaps 2;  
QY 31 PGAAAHLPWWPPQLAASRREAPPLSORPHT-----OGAGSPPETNEKLTN 78  
DB 2392 PNSAPAHLP LNPP-----RIQPIAQLPIKCTPACTVSNANPQSGPPRVEFDNN 2444  
QY 79 P 79  
DB 2445 P 2445

Search completed: November 26, 2002, 19:44:26  
Job time : 46.2581 secs